

N 69 14251

NASA CR 98653

INVESTIGATION OF CONTINENTAL DRIFT

Phase-I Effort

**CASE FILE
COPY**

Contract NSR 09-015-079

1 April 1968 through 30 September 1968

November 1968

Coinvestigators: Mr. Carlton G. Lehr
Dr. Ursula B. Marvin
Dr. Paul A. Mohr

Prepared for
National Aeronautics and Space Administration
Washington, D. C. 20546

Smithsonian Institution
Astrophysical Observatory
Cambridge, Massachusetts 02138

INVESTIGATION OF CONTINENTAL DRIFT

Phase-I Effort

Contract NSR 09-015-079

1 April 1968 through 30 September 1968

November 1968

Coinvestigators: Mr. Carlton G. Lehr
Dr. Ursula B. Marvin
Dr. Paul A. Mohr

Prepared for
National Aeronautics and Space Administration
Washington, D. C. 20546

Smithsonian Institution
Astrophysical Observatory
Cambridge, Massachusetts 02138

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
1 SUMMARY	1
2 GEOLOGICAL CONSIDERATIONS	6
3 INTERFEROMETER SYSTEMS	18
3.1 Basic Principles of Interferometry	18
3.2 Independent-Clock Interferometers	25
3.3 Use of an Artificial Source	38
4 LASER SYSTEMS	43
4.1 Present Laser Systems	43
4.2 Possible Improvements in Laser Systems	50
5 ATMOSPHERIC AND IONOSPHERIC INFLUENCES	65
5.1 Introduction	65
5.2 Atmospheric and Ionospheric Differential-Phase Path Length and Its Fluctuations	65
5.3 Columnar-Refractivity Measurements in the Direction of Observation	74
5.4 Conclusions and Recommendations	76
6 MODELING THE SYSTEM	78
7 REFERENCES	84

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1	Block diagram illustrating schematically the configurations and roles of the lithosphere, asthenosphere, and mesosphere in a version of the new global tectonics in which the lithosphere, a layer of strength, plays a key role	6
2	The locations of the boundaries of the six blocks used in the computations	7
3	Available data on sea-floor spreading	9
4	Summary map of slip vectors derived from earthquake-mechanism studies	11
5	Worldwide distribution of all earthquake epicenters for the period 1961 through 1967	12
6	Structural features of the Gulf of California	15
7	Structural features of the Gulf of Aden	16
8	Basic geometrical relationships in interferometry	18
9	Interferometer outputs: (a) adding interferometer; (b) phase-switched interferometer.	20
10	Sample fringe amplitude versus frequency for different relative frequency offsets: (a) 0.0 Hz; (b) 0.025 Hz; (c) 0.05 Hz	32
11	Base-line-satellite geometries for various aspects of satellite interferometry.	41
12	Diagram for threshold detection near the point of maximum slope	54
13	Spherically symmetric model of dry-air atmosphere	67
14	Ionospheric electron-density profile	67
15	Long-term changes in electrical path length and refractivity, Maui path (November 6-9, 1956)	69
16	Spectrum of apparent path-length variations, Maui path (November 6-9, 1956)	70
17	Long-term changes in electrical path length and refractivity, Green Mountain Mesa-Table Mesa path (September 15-17, 1958)	71
18	Long-term changes in electrical path length and refractivity, Green Mountain Mesa-Table Mesa path (October 10-15, 1958)	72

<u>Figure</u>		<u>Page</u>
19	Spectrum of apparent path-length variations, Green Mountain Mesa—Table Mesa path (October 10-15, 1958)	73
20	Proposed radar arrangement for the monostatic measurement of the atmospheric and ionospheric columnar refractivity in the wanted direction of observation	76
21	Preliminary flow chart for total position measurement system	81

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Instantaneous centers of rotation deduced from Table 2 of Le Pichon (1968)	10
2	Measured spreading rates	10
3	Azimuths of fracture zones used for least-squares fitting . . .	10
4	Computed differential movements between blocks as given in Figure 6 from Le Pichon (1968)	13
5	Relative crustal motions of Baker-Nunn station sites, calculated from continental-drift models of Morgan (1968) and Le Pichon (1968)	17
6	Satellites with retroreflectors	45
7	Characteristics of geodetic laser-ranging systems	47
8	Surface-to-free-space differential-phase path length	66

INVESTIGATION OF CONTINENTAL DRIFT

Phase-I Effort

Contract NSR 09-015-079

1. SUMMARY

S. Ross

This study was undertaken to examine the feasibility of applying two promising techniques of high-precision distance measurement to test for the existence of continental drift. The techniques involved are ultrashort-pulse laser ranging and independent-clock radio interferometry. Specifically, the main goals of the study were 1) to investigate criteria for determining likely geophysical areas for the measurement of continental drift, 2) to review the principles of operation and to delineate the relative advantages of the two techniques, and 3) to formulate a preliminary model of the overall system that could eventually be used for the comparative evaluation of candidate observation programs.

Marvin's survey (1966) of continental drift served as the starting point for the study of likely geophysical measurement areas. We found that during the 2 years since the publication of this article, the geophysics community has shown an unparalleled increase of interest in the related phenomena of continental drift, sea-floor spreading, and global tectonics. These findings are summarized in Section 2. Perhaps the most significant of the major recent developments has been the formulation of a more comprehensive theory of gross crustal motion; the theory subdivides the earth's surface into six major blocks, each undergoing slow rotational motion relative to the others. This has given rise to some preference for the use of the phrase "crustal block motion" rather than "continental drift" to describe the phenomenon; however, it has been convenient in the context of this study to use the two terms interchangeably.

From examination of geomagnetic and geophysical evidence, Le Pichon (1968) and Morgan (1968), in separate studies, located the poles of rotation and gave estimates for the magnitudes of the rotation vectors at these poles. It is encouraging that their determinations of these quantities are in good general agreement. From their data it is possible to estimate the relative drift rates of various pairs of points on the earth's surface; as an example, we have tabulated in Section 2 the relative motions between pairs of stations in the Baker-Nunn network, as predicted by the drift theory.

Since the first successful demonstration of the independent-clock interferometer technique some 18 months ago, the number of interferometer experiments conducted by various groups has been increasing rapidly. Baseline distances have been steadily lengthened so that now measurements have been made between such widely separated points as Green Bank, West Virginia, and Onsala, Sweden; and Canberra, Australia, and Goldstone, California.

Our study of radio interferometry has familiarized us with the operational aspects of recording and reducing radio signals received by such interferometers. Fringe-processing software obtained from MIT Lincoln Laboratory and from the National Radio Astronomy Observatory (NRAO) was studied; portions of the former were converted to the CDC 6400 computer at Smithsonian Astrophysical Observatory (SAO). We had two principal objectives in making this conversion: The first was to achieve, through step-by-step tracing of the instruction code, a fundamental understanding of the details of the technique so that we could judge its advantages and limitations for the stated goals of the study. The second was to acquire the ability of processing raw interferometry data if this should prove necessary. Section 3 summarizes the theory of the interferometer as it is applied to accurate positional determinations and, within this framework, outlines the steps that enter into the reduction of the data.

Laser satellite-ranging became practical several years ago. Nevertheless, laser systems are recent enough so that substantial evolutionary improvement occurs from year to year. Thus, existing satellite-tracking systems have an inherent accuracy of 1 m; the SAO system at Mt. Hopkins is an example of such a system. However, the way seems clear for achieving accuracies approaching a few centimeters, which brings the technique into competition as a method for conducting continental-drift measurements. Section 4 reviews the current status of laser systems and the prospects for improving them.

Our study identified atmospheric and ionospheric turbulence and unpredictability as the principal contributors to signal distortion (and therefore positional errors) in both the radio and the optical approaches. We conducted an extensive survey of pertinent literature, including a review of the refractive effects of ionized and neutral atmospheres upon microwave propagations. Our findings are presented in Section 5, along with a preliminary assessment of proposed techniques for lessening the uncertainties in the interferometer and laser measurements. The techniques investigated included analytical approaches, such as ray-tracing methods applied to atmospheric models as a means of removing refractive bias, and instrumental approaches using radiosonde and refractometer probings, two-frequency measurements of ionospheric and atmospheric columnar refractivity, and laser and microwave measurements of incoherent backscatter from ionospheric electrons and atmospheric molecules.

For the establishment of an observing program to test for continental drift, we considered a number of possible approaches. If a laser system is used, the corner reflector can be placed on an artificial satellite or on the moon. An interferometer, besides being able to view a stellar radio source, can also make use of a transmitter placed on a natural or an artificial satellite. The observational program can employ either a geometrical or a dynamical approach, or some combination of them, irrespective of the type of instrument used. The geometrical approach employs simultaneous

measurements of the same source from two or more stations, while in the dynamical approach, the source is observed by different stations at different times; the latter requires that a very precise orbit be determined to describe the motion of the observed object.

When derived geometrically, both the laser and the interferometer results are insensitive to translations of the entire network of stations. However, the stellar interferometer has the advantage of being able to sense rigid rotations of the total configuration since the source positions enter into the analysis as absolute spatial directions against which the rotations can be referenced. Sensing such rotations is important in distinguishing true continental drift from polar wandering; in the latter, the entire crust moves rigidly with respect to the earth's angular-momentum vector.

The dynamical approach is limited in its accuracy by the precision with which the satellite orbit can be determined, or by the lunar ephemeris and librations. In the dynamical approach, the center of mass of the earth is used as a reference point and enters into the solution through the orbital equations that describe the motion of the satellite or of the moon around it. Thus, it is possible to determine the absolute positions of the individual stations with respect to the center of mass, and from this to detect translational motions of the station network, which, as noted above, would be impossible with the geometric approach.

The laser method is based on range-only measurements from each station to a corner reflector. In contrast, the interferometer responds to range differences to the source, from which (Section 3) it is possible to calculate both the base-line direction and the base-line length. Network configurations derived from triangulations that use distance measurements alone are more generally subject to error propagation than are methods that employ directional information in combination with the range data. This fact argues for an approach that includes both range and directional measurements.

From an operational standpoint, the laser system is less expensive to install and operate, it is much more readily transportable, and the procedures involved in processing its measurements are simpler and faster to perform. However, the laser system cannot be used in the presence of clouds or too bright an ambient sunlight background, although there are some indications that certain propagation effects may be less pronounced at optical than at radio frequencies.

All these considerations make it clear that any realistic program for measuring continental drift will have to include a variety of coordinated measurement techniques and a combination of observational approaches, with artificial satellites certain to play a leading part. The interplay of factors favoring one or the other method strongly suggests that the development of a simulation tool for modeling the candidate systems might be well advised. Such a tool, which could be used for the comparative evaluation of promising alternatives, is described in Section 6, along with a preliminary description of the functional structure of a model for conducting the systems simulations.

2. GEOLOGICAL CONSIDERATIONS

P. A. Mohr

In the short period since initiation of this study, there has been a remarkable advance in knowledge of the physical nature of continental drift. This knowledge largely stems from the successful synthesis of the global tectonic pattern, using the models established by Morgan (1968) and Le Pichon (1968). These models essentially consider the earth's crust (the lithosphere) as composed of a number of rigid plates (or "paving stones") free to move over a plastic or fluid upper mantle (the asthenosphere; see Figure 1). These crustal plates may be composed entirely of oceanic or continental lithosphere or of a coupled ocean-continent lithosphere (Figure 2).

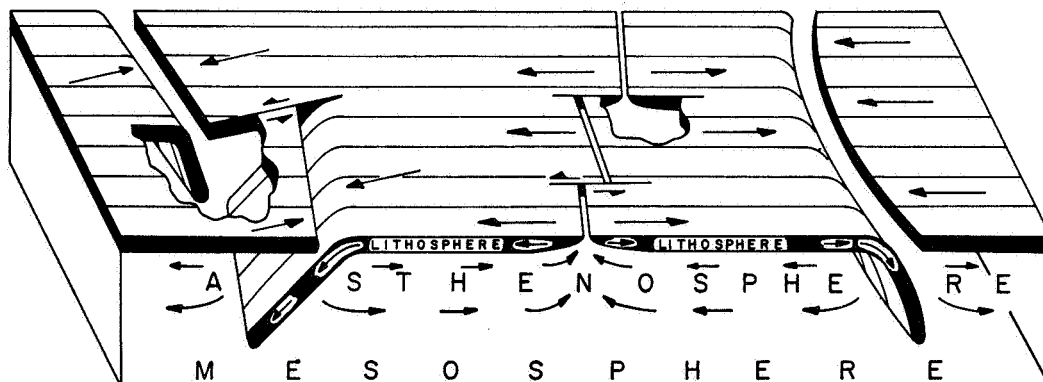


Figure 1. Block diagram illustrating schematically the configurations and roles of the lithosphere, asthenosphere, and mesosphere in a version of the new global tectonics in which the lithosphere, a layer of strength, plays a key role. Arrows on lithosphere indicate relative movements of adjoining blocks. Arrows in asthenosphere represent possible compensating flow in response to downward movement of segments of lithosphere. One arc-to-arc transform fault appears at left between oppositely facing zones of convergence (island arcs), two ridge-to-ridge transform faults along ocean ridge at center, simple arc structure at right. From Isacks, Oliver, and Sykes (1968).

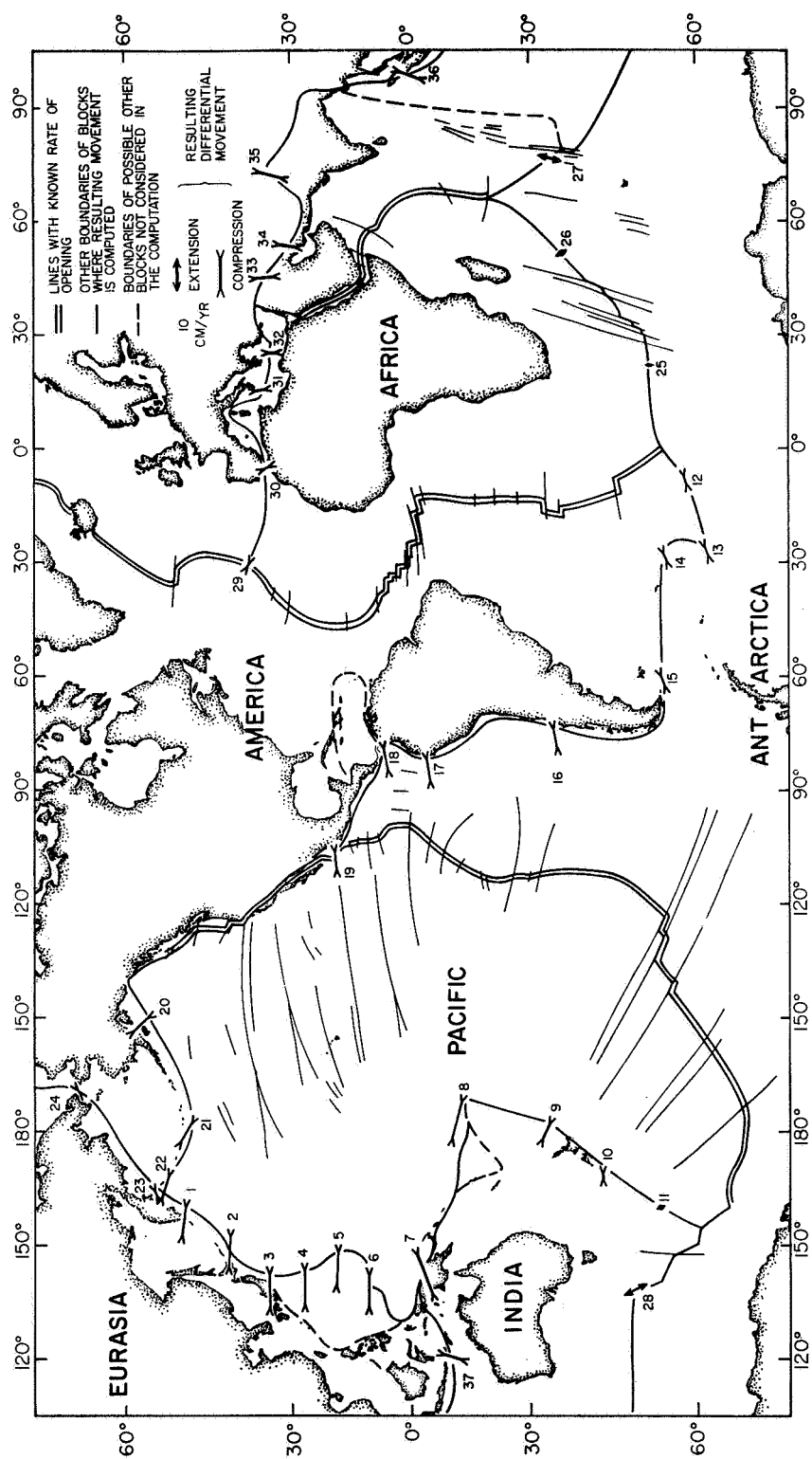


Figure 2. The locations of the boundaries of the six blocks used in the computations. The numbers next to the vectors of differential movement refer to Table 4. Note that the boundaries where the rate of shortening or slippage exceeds about 2 cm yr^{-1} account for most of the world earthquake activity. From Le Pichon (1968).

The midocean ridge rifts mark lines where two crustal plates are moving apart, in response to dilation and formation of new basaltic crust at the ridge rift itself. The island arcs and the related fold mountains mark lines where two plates converge and one overrides the other (see Figure 1; note that either plate may override the other, not solely the plate adjoining the mid-ocean ridge rift).

Morgan and Le Pichon show that the global pattern of oceanic ridge rifts and island arcs is compatible with the rotational motion of a discrete number of large lithospheric plates (Figure 3; Table 1). Furthermore, by use of known data on the rates of sea-floor spreading, obtained from the spacings of magnetic lineations of known age (Table 2), and on the orientation of spreading, obtained from the azimuths of transform faults associated with an offset spreading pattern (Table 3), an excellent integration with the global tectonic pattern is obtained.

Recent seismological studies (Isacks et al., 1968) strongly confirm the Morgan-Le Pichon theory (Figures 4 and 5). It is therefore opportune to take the data of Le Pichon (Figure 3; Table 4) as a basis for calculating the expected rate and direction of movement between any two points on the earth's surface.

Table 4 shows that the most rapid horizontal crustal movements (up to 10 cm yr^{-1}) are to be expected in the island arcs; this conclusion is in accord with the observation that the greater number and highest magnitude of earthquakes occur along these compressional lines (Figure 5). On the other hand, the oceanic ridge rifts, where crustal dilation is occurring (although often on a smaller scale than island-arc compression), have a better understood mechanism and a much more precisely established crustal-movement rate, based on the factors mentioned above, than do the island arcs.

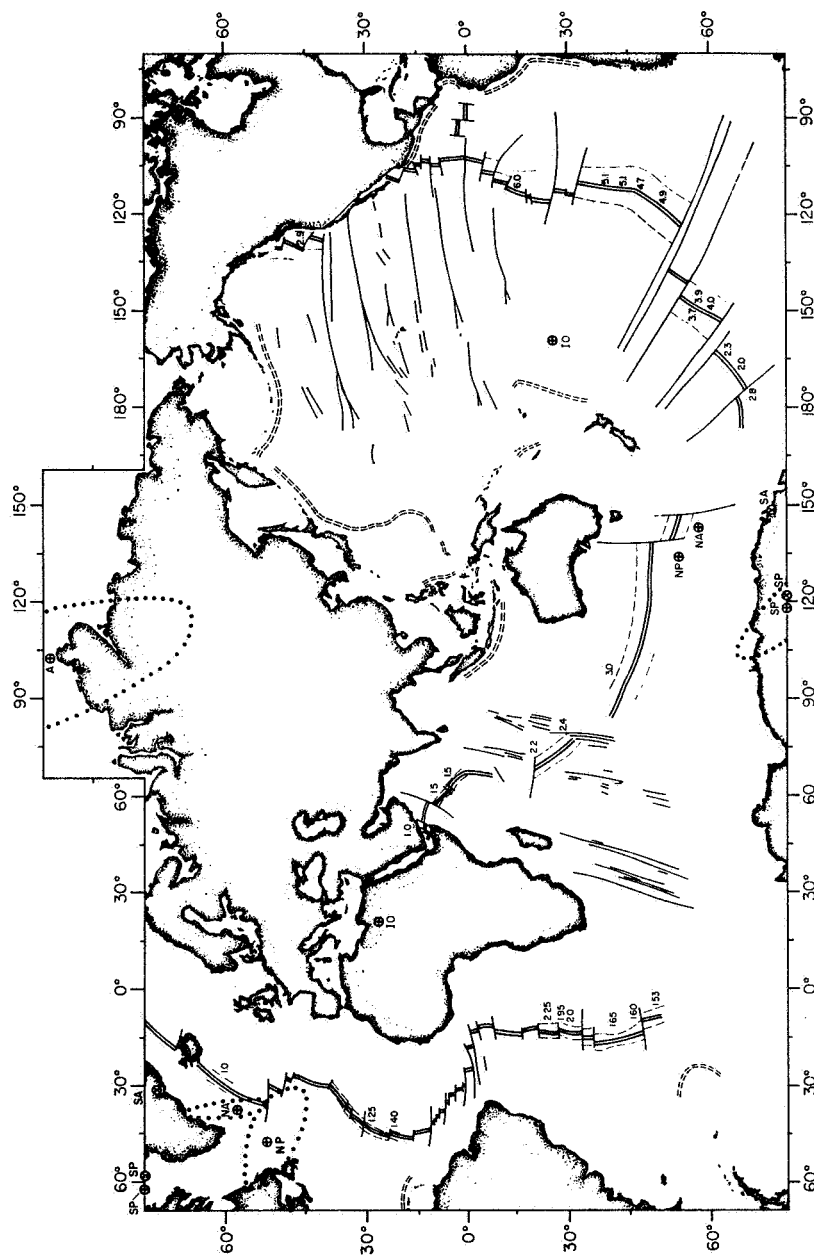


Figure 3. Available data on sea-floor spreading. The axes of the actively spreading midocean ridges are shown by a double line; the fracture zones, by a single line; anomaly 5 (≈ 10 m. y. old), by a single dashed line; the active trenches, by a double dashed line. The spreading rates are given in centimeters per year. The locations of the centers of rotation obtained from spreading rates are shown by X; those obtained from the azimuths of the fracture zones by +. NA stands for North Atlantic; SA for South Atlantic; NP for North Pacific; SP for South Pacific; IO for Indian Ocean; A for Arctic. The ellipses drawn around the NA, NP, SP, and A centers of rotation obtained from the fracture zones are the approximate loci of the points at which the standard deviation equals 1.25 times the minimum standard deviation. The ellipse around the IO center of rotation is too small to be shown. These ellipses indicate how fast the least-squares determination converges. From Le Pichon (1968).

Table 1. Instantaneous centers of rotation deduced from Table 2 of Le Pichon (1968). From Le Pichon (1968).

	Latitude	Longitude	Rate,* 10 ⁻⁷ deg/yr
America-Antarctica	79.9S	40.4E	-5.44
Africa-Antarctica	42.2S	13.7W	+3.24
India-Antarctica	4.5S	18.1E	+5.96
India-Eurasia	23.0N	5.2W	-5.50
India-Pacific	52.2S	169.2E	-12.3
Eurasia-Pacific	67.5S	138.5E	-8.15
Africa-Eurasia	9.3S	46.0W	-2.46

* Positive value indicates extension; negative, compression.

Table 2. Measured spreading rates.* From Le Pichon (1968).

Pacific			Atlantic			Indian		
Latitude	Longitude	Rate, cm/yr	Latitude	Longitude	Rate, cm/yr	Latitude	Longitude	Rate, cm/yr
48N	127W	2.9	60N	29W	0.95	19N	40E	1.0
17S	113W	6.0 (5.9†)	28N	44W	1.25 (1.3†)	13N	50E	1.0
40S	112W	5.1 (5.3)	22N	45W	1.4 (1.5)	7N	60E	1.5
45S	112W	5.1 (5.1)	25S	13W	2.25 (2.0)	5N	62E	1.5
48S	113W	4.7 (5.0)	28S	13W	1.95 (2.0)	22S	69E	2.2
51S	117W	4.9 (4.8)	30S	14W	2.0 (2.0)	30S	76E	2.4
58S	149W	3.9 (3.6)	38S	17W	2.0 (1.9)	43S	93E	3.0
58S	149W	3.7 (3.6)	41S	18W	1.65 (1.9)			
60S	150W	4.0 (3.4)	47S	14W	1.60 (1.6)			
63S	167W	2.3 (2.8)	50S	8W	1.53 (1.5)			
65S	170W	2.0 (2.6)						
65S	174W	2.8 (2.4)						

* Arctic Ocean: ≈ 1.0 cm/yr.

Norwegian Sea: ≈ 1.0 cm/yr.

† Computed from center of rotation determined from spreading rates by least squares.

Table 3. Azimuths of fracture zones used for least-squares fitting.* From Le Pichon (1968).

South Pacific			Indian Ocean			Arctic Ocean		
Latitude	Longitude	Azimuth	Latitude	Longitude	Azimuth	Latitude	Longitude	Azimuth
56S	122W	117 (114†)	12N	46E	35 (34†)	82N	12W	132 (131†)
56S	124W	118 (115)	13N	50E	30 (30)	71.5N	12W	125 (116)
57S	141W	121 (121)	9N	55E	32 (32)	66.5N	20W	98 (110)
62S	152W	123 (128)	15N	60E	24 (23)	52N	35W	96 (101)
63S	160W	125 (131)	18N	61E	19 (19)			
65S	170W	141 (137)						

* Azimuths of fracture zones for the Atlantic Ocean and the North Pacific Ocean are as listed by Morgan [1968].

† Computed from center of rotation determined from fracture zones by least squares.

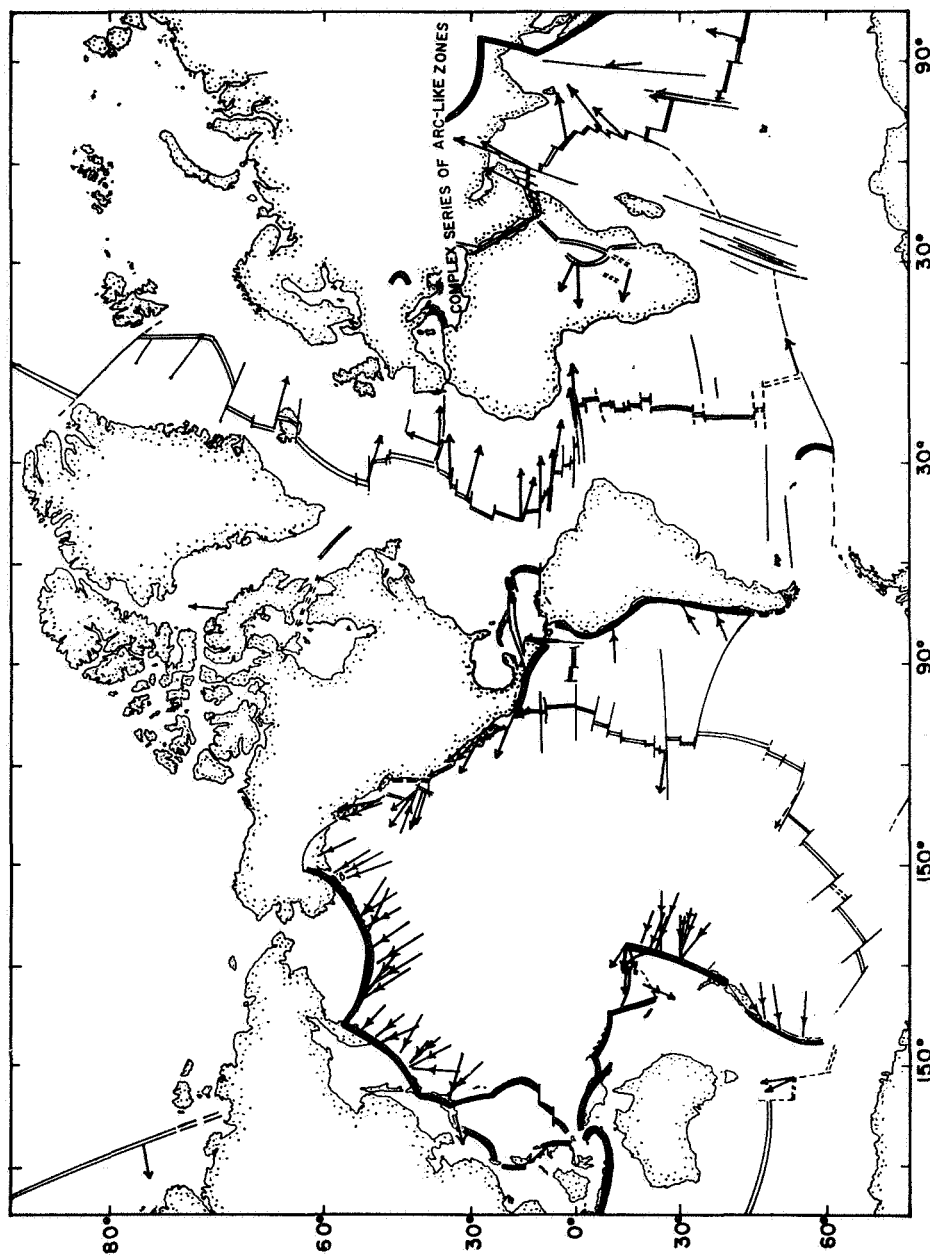


Figure 4. Summary map of slip vectors derived from earthquake-mechanism studies. Arrows indicate horizontal component of direction of relative motion of block on which arrow is drawn to adjoining block. Crests of world rift system are denoted by double lines; island arcs and arc-like features, by bold single lines; major transform faults, by thin single lines. Both slip vectors are shown for earthquakes near the western end of the Azores-Gibraltar ridge since a rational choice between the two could not be made. Compare with directions computed by Le Pichon (1968, Figure 2). From Isacks et al. (1968).

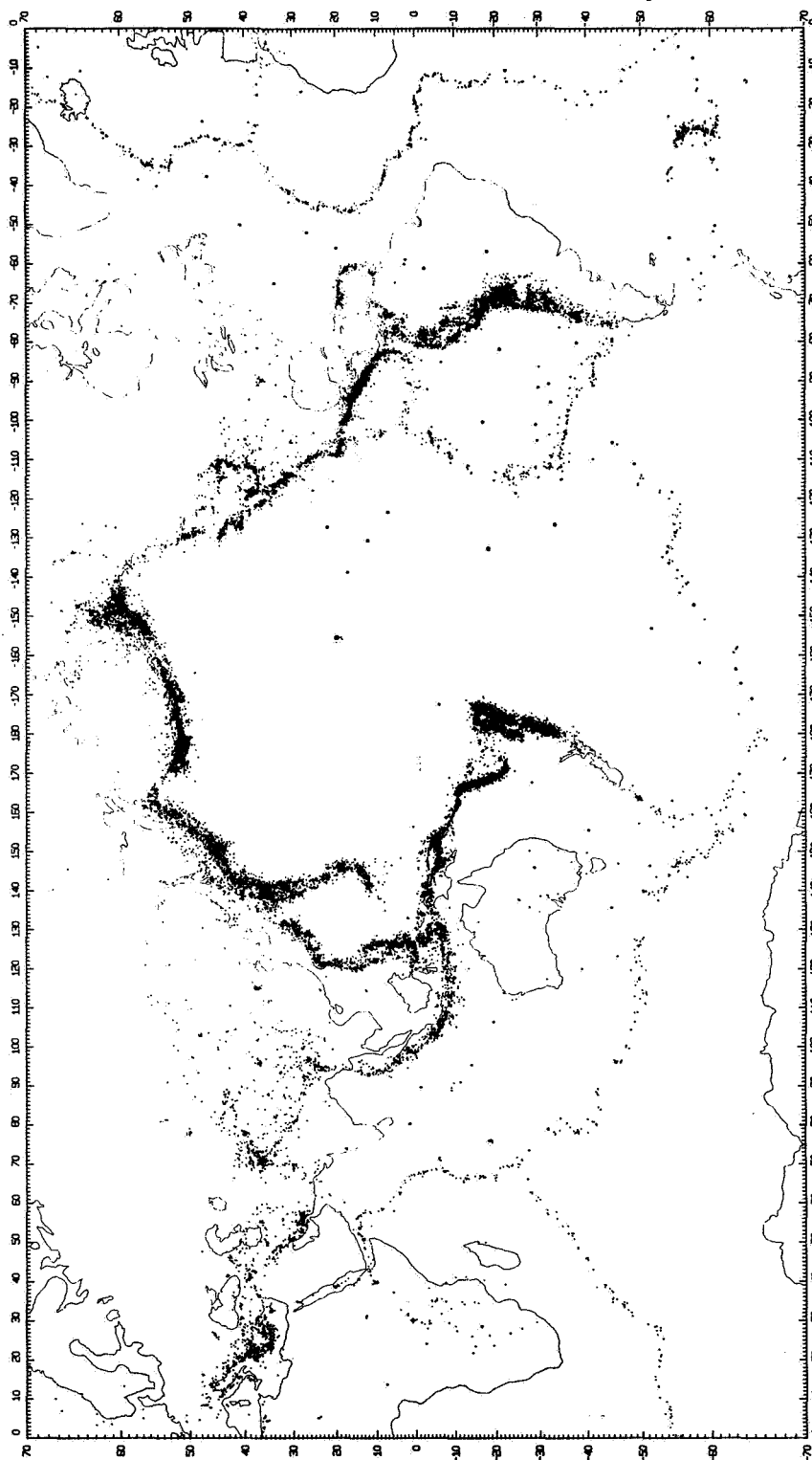


Figure 5. Worldwide distribution of all earthquake epicenters for the period 1961 through 1967 as reported by the U.S. Coast and Geodetic Survey (after Barazangi and Dorman, 1968). Note continuous narrow major seismic belts that outline aseismic blocks; very narrow, sometimes step-like pattern of belts of only moderate activity along zones of spreading; broader, very active belts along zones of convergence; diffuse pattern of moderate activity in certain continental zones. From Isacks et al. (1968).

Table 4. Computed differential movements between blocks as given in Figure 6 from Le Pichon (1968). From Le Pichon (1968).

	Latitude	Longitude	Rate,* cm/yr	Azimuth	Location
Eurasia-Pacific					
1.	51N	160E	-7.9	99	Kurile Trench
2.	43N	148E	-8.5	93	Kurile Trench
3.	35N	142E	-8.8	91	Japan Trench
4.	27N	143E	-9.0	91	Japan Trench
5.	19N	148E	-9.0	93	Mariana Trench
6.	11N	142E	-8.9	91	Mariana Trench
India-Pacific					
7.	3S	142E	-11.0	69	New Guinea
8.	13S	172W	-9.1	107	N. Tonga Trench
9.	34S	178W	-4.7	113	S. Kermadec Trench
10.	45S	169E	-1.7	89	S. New Zealand
11.	55S	159E	+1.6	339	Macquarie Island
America-Antarctica					
12.	58S	7W	-2.6	72	Southwest Atlantic
13.	61S	26W	-2.7	68	S. South Sandwich Trench
14.	55S	29W	-3.3	72	N. South Sandwich Trench
15.	55S	60W	-3.7	73	Cape Horn
16.	35S	74W	-5.2	79	S. Chile Trench
17.	4S	82W	-6.0	81	N. Peru Trench
18.	7N	79W	-5.9	80	Panama Gulf
19.	20N	106W	-5.3	83	N. Middle America Trench
America-Pacific					
20.	57N	150W	-5.3	136	E. Aleutian Trench
21.	50N	178W	-6.2	119	W. Aleutian Trench
22.	54N	162E	-6.3	108	W. Aleutian Trench
America-Eurasia					
23.	56N	165E	-1.6	68	Aleutian-Kurile Islands
24.	66N	169W	-1.4	62	Alaska-Siberia
Africa-Antarctica					
25.	53S	22E	+1.5	8	S. Southwest Indian Ridge
26.	37S	52W	+2.7	331	N. Southwest Indian Ridge
India-Antarctica					
27.	36S	75E	+5.8	17	W. Southeast Indian Ridge
28.	50S	138E	+6.4	333	E. Southeast Indian Ridge
Africa-Eurasia					
29.	40N	31W	-1.5	117	Azores
30.	36N	6W	-1.9	153	Gibraltar
31.	38N	15E	-2.4	169	Sicily
32.	35N	25E	-2.6	176	Crete
India-Eurasia					
33.	37N	45E	-4.3	176	Turkey
34.	30N	53E	-4.8	186	Iran
35.	35N	72E	-5.6	193	Tibet
36.	0	97E	-6.0	23	W. Java Trench
37.	12S	120E	-4.9	19	E. Java Trench

* Positive value indicates extension; negative, compression.

Thus, it would be felicitous to check the new global tectonic theory observations of where an oceanic ridge rift enters a land mass, thereby providing relatively closely spaced shores to either side of the spreading line. Observing stations placed near each shore could then be geodetically tied together, with consequent less risk of local intervening tectonics and intracrustal plate deformation than is the case over continental distances. Two localities suggest themselves: 1) the Gulf of California (Figure 6), and 2) the Gulf of Aden-Red Sea (Figure 7).

Table 5 lists, on the basis of the Morgan-Le Pichon hypothesis, the relative movements of existing Baker-Nunn satellite-tracking stations. While use of these stations in continental-drift studies would obviously be desirable, especially because they will soon be equipped with lasers, their sites were not originally chosen with this program in mind. It would therefore be useful to make a computer study (allied to a consideration of weather and political factors) to find the best locations of maximum relative drift.

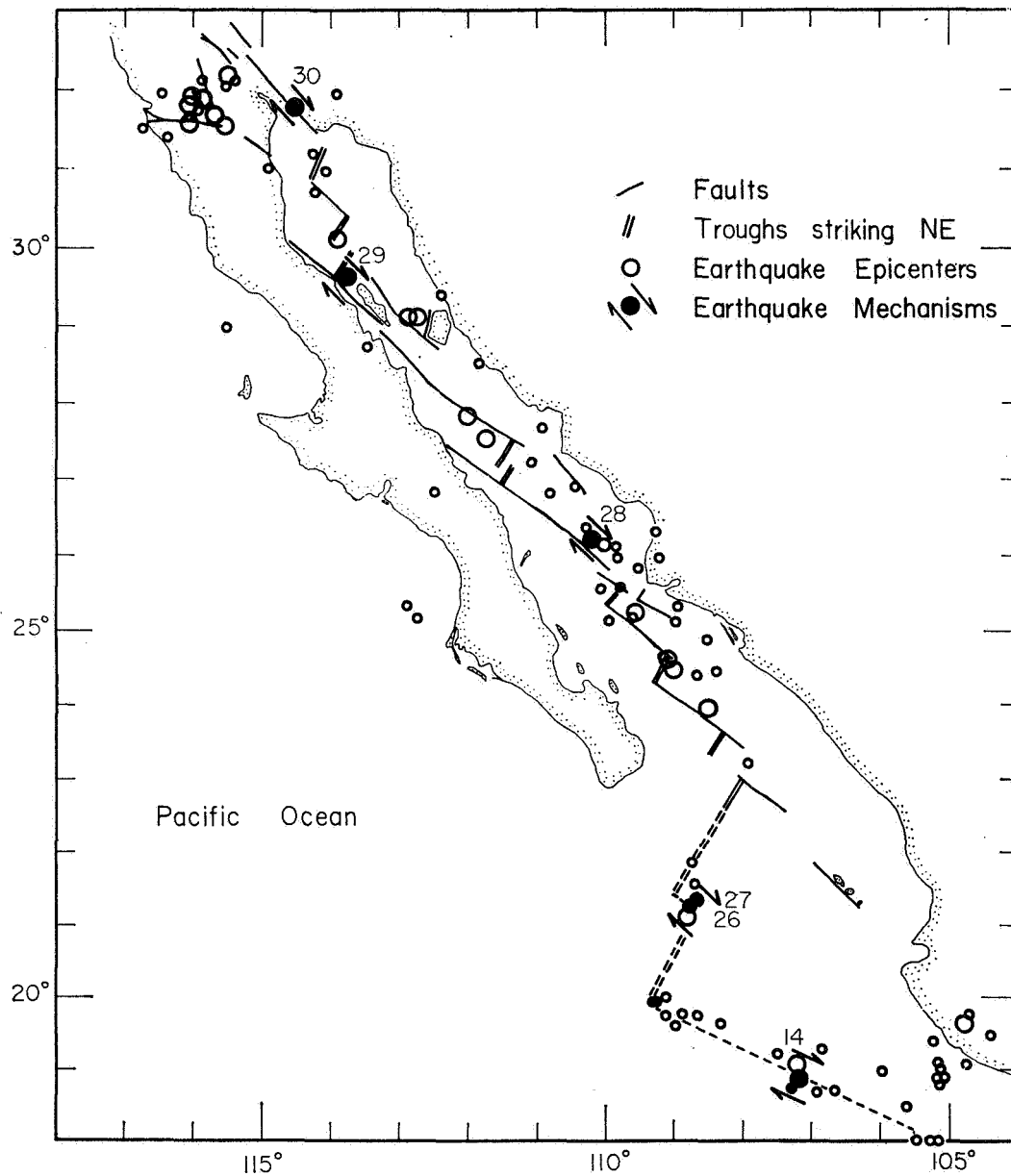


Figure 6. Structural features of the Gulf of California (after Sykes, 1968). Relocated epicenters of earthquakes for the period 1954 to 1962. Seismicity and focal mechanisms support the hypothesis of spreading by ocean-ridge-transform-fault mechanism. From Isacks et al. (1968).

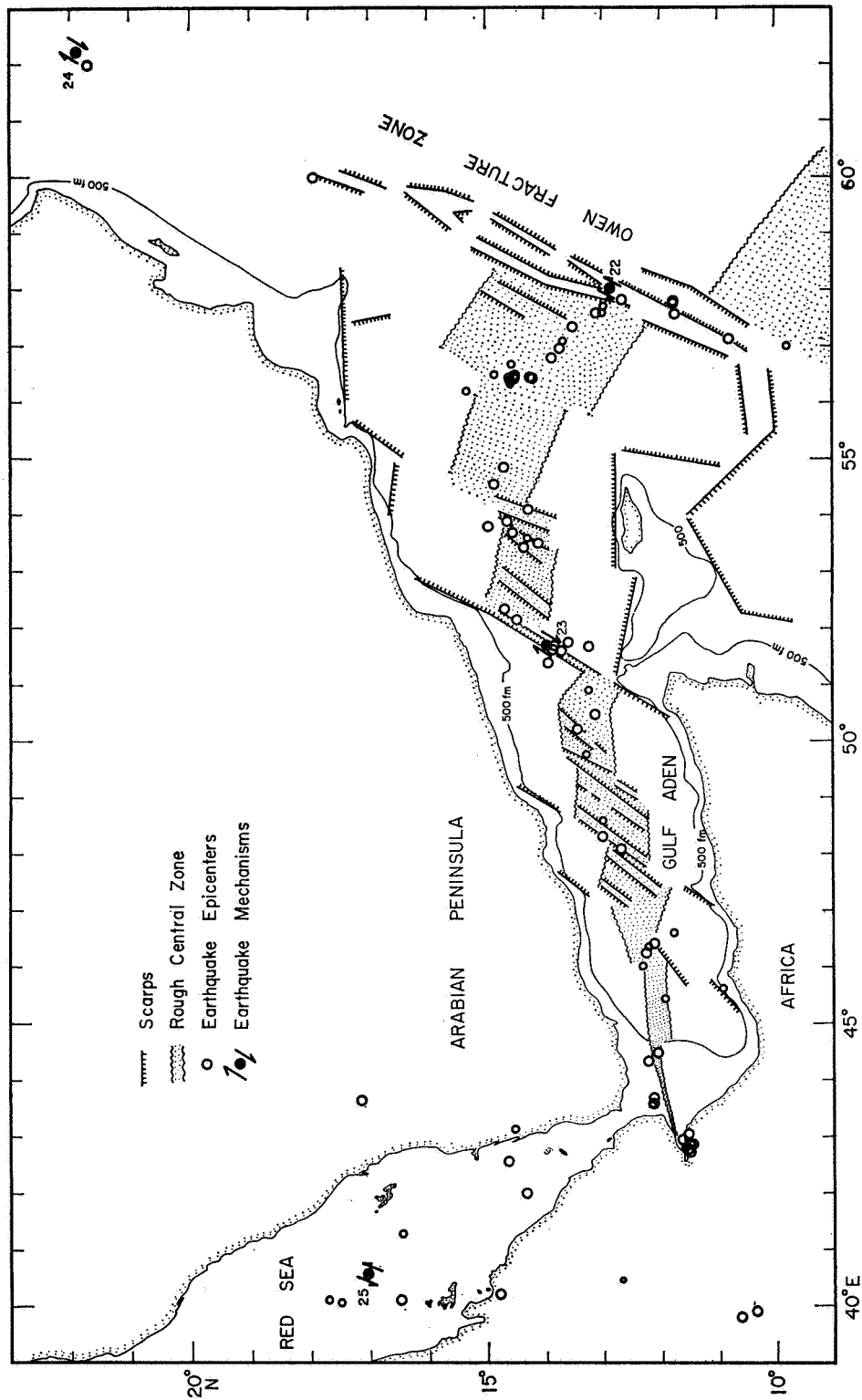


Figure 7. Structural features of the Gulf of Aden (after Sykes, 1968). Relocated epicenters of earthquakes for the period 1955 to 1966. Scarps and rough central zone after Laughton (1966). Seismicity and focal mechanisms support hypothesis of spreading by ocean-ridge-transform-fault mechanism. From Isacks et al. (1968).

Table 5. Relative crustal motions of Baker-Nunn station sites, calculated from continental-drift models of Morgan (1968) and Le Pichon (1968)

Station abbreviations:

OL - Olifantsfontein	MH - Mt. Hopkins	CR - Comodoro Rivadavia
SF - San Fernando	IL - Island Lagoon	AT - Athens
NT - Naini Tal	DO - Dodaira	RO - Rosamond
AR - Arequipa	DZ - Debre Zeit	CL - Cold Lake
MI - Maui	NA - Natal	JI - Johnston Island

No relative motions between:

MI and JI	DZ and OL
CL, RO, and MH (?)	NT and IL (?)
AR, NA, and CR (?)	AT and SF (?)

Station pairs *	Motion (cm yr ⁻¹)	Azimuth of relative motion (degree)
MI - CL	6	100
-RO (MH)	6	120
-AR	5	120
-DO	9	270
JI - RO (MH)	9	120
-DO	9	270
-IL	10	90
BI - SF	2.5	160
-AT	2.5	180
-NT	2.5	40
-NA	5	260
OL - SF	2.5	160
-AT	2.5	180
-NT	3	50
-NA	4.5	270
SF - NA	3.5	270
AT - NT	4	000
NT - DO	6	200
-JI	10	260
IL - DO	5	200

*The reference station is given first.

3. INTERFEROMETER SYSTEMS

S. Ross

3.1 Basic Principles of Interferometry

In simplest terms, an interferometer is an apparatus for measuring the phase difference between simultaneously received electromagnetic waves that emanate from an infinitely distant source. Given ω , the nominal frequency of the signal, then $\Delta\phi$, the phase difference between the signals received simultaneously at any two terminals, can be used to determine the elapsed time between the arrival of any particular wave front at each of the two in turn, viz., $\Delta t = \Delta\phi/\omega$. And if we know either the phase difference or the elapsed time, we can accurately express the source direction in terms of the base-line vector connecting the two terminals, or vice versa, the latter case being of more direct interest in the present context. Simple geometry (Figure 8) shows that $\Delta\phi = (D/\lambda) \cos \theta$, which relates the base-line length to the source direction, where λ is the signal wavelength and $\Delta\phi$ is measured in cycles.

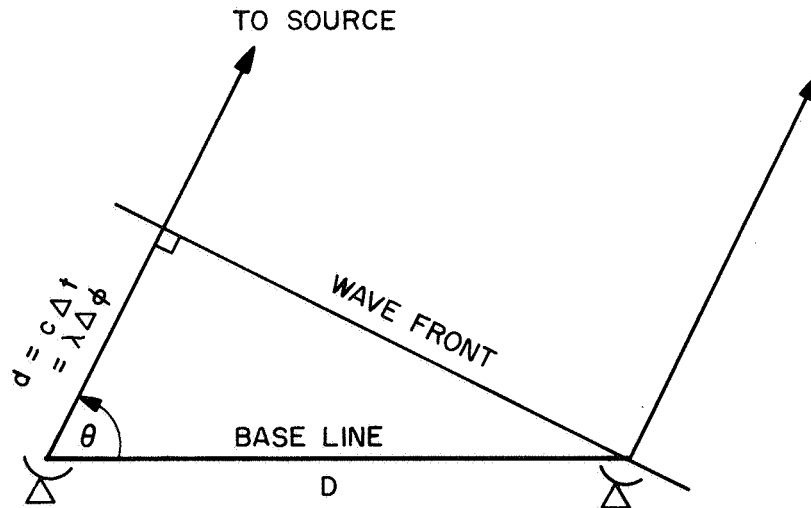


Figure 8. Basic geometrical relationships in interferometry.

When the value of θ is such that $\Delta\phi$ is an integral number of half-cycles (i. e., $\theta = \cos^{-1} (n\lambda/2D)$, where n is an integer), the signals received at each terminal are in phase or antiphase, and a relative extremum of power is available from the interferometer.

Let us assume for simplicity that the signal received at one antenna is of the form $e_1(t) = A_1(\theta) \cos \omega t$, and at the second antenna, $e_2(t) = A_2(\theta) \cos (\omega t + \Delta\phi)$, where $A_1(\theta)$ and $A_2(\theta)$ are the individual antenna-response characteristics, i. e., their far-field patterns. For the simple (or adding) interferometer, the combined signal* will be

$$e(t) = e_1(t) + e_2(t) = A_1 \cos \omega t + A_2 \cos (\omega t + \Delta\phi) \quad .$$

The available power from the interferometer can then be expressed as

$$\begin{aligned} P(\theta) &= A_1^2 \cos^2 \omega t + A_2^2 \cos^2 (\omega t + \Delta\phi) + 2 A_1 A_2 \cos \omega t \cos (\omega t + \Delta\phi), \\ &= \frac{1}{2} (A_1^2 + A_2^2) + A_1 A_2 \cos \Delta\phi + \text{terms at radio frequency } \dots, \end{aligned}$$

where the radio-frequency terms make no integrated contribution to the power over the sample interval.

If the antennas have identical characteristics, $P(\theta)$ becomes

$$P(\theta) = |A(\theta)|^2 (1 + \cos \Delta\phi),$$

which has the form of a sinusoid modulated by the far-field pattern response of a single antenna (Figure 9a). The sinusoidal response pattern is termed the interference fringes, by analogy with the characteristic pattern of

* Disregarding, for the present, all effects due to the receiving and processing equipment.

concentric fringe bands formed by an optical interferometer. When the source is near the zenith, $\theta \approx \pi/2$, and the period of the sinusoid approaches λ/D .

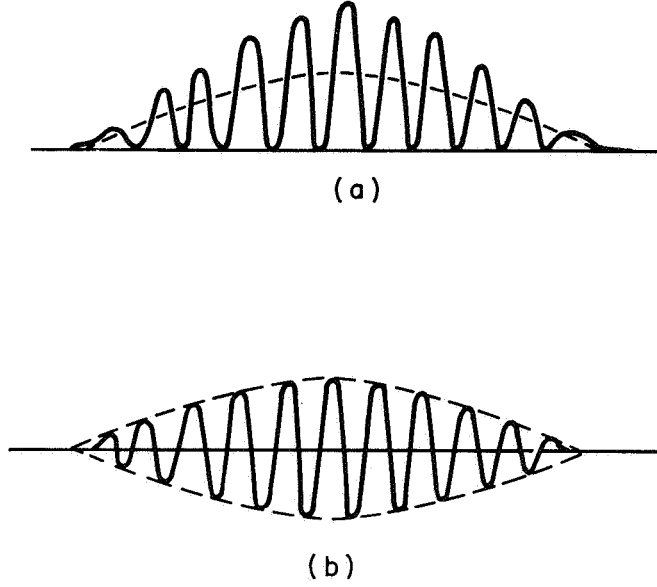


Figure 9. Interferometer outputs: (a) adding interferometer; (b) phase-switched interferometer.

Let us suppose a second source is located near the first. Then the interferometer's response to it will be

$$P'(\theta) = |A(\theta)|^2 (1 + \Delta\phi') \quad ,$$

and the combined response to both sources will be

$$P_T = P + P' = 2 A^2 \left[1 + \cos\left(\frac{\Delta\phi_1 + \Delta\phi_2}{2}\right) \cos\left(\frac{\Delta\phi_1 - \Delta\phi_2}{2}\right) \right] \quad .$$

If the two sources are brought very close together, say to within δ rad of one another, then P_T becomes

$$P_T \approx 2 A^2 \left[1 + \cos \Delta\phi \cos \left(\frac{D\delta}{\lambda} \sin \theta \right) \right] .$$

When $(D\delta/\lambda) \sin \theta = \pi/2$, i. e., when $\delta = \lambda\pi/(2D \sin \theta)$, the fringes disappear and the source is said to be resolved. When the interferometer is operated near $\theta = \pi/2$, the resolution criterion becomes $\delta = \lambda\pi/2D$, which indicates that the resolution of the interferometer near $\theta = \pi/2$ is twice that of a single aperture whose width is equal to the base-line length.

As the source transits the interferometer, the power (or intensity) response shown in Figure 9a is produced; it can be observed, for example, by driving a strip recorder from the amplifier. In such a case, the abscissa is time, which of course is linearly related to θ . Now, if the time at which a specific fringe is produced can be determined precisely enough, the relationship cited earlier for $\Delta\phi$ in terms of source position and base-line parameters can be equally precisely specified. In long-base-line measurements, however, the fringe density is so great that it is very difficult to identify the central fringe* and hence very difficult to record accurately the time of passage through any n^{th} -order fringe (i. e., the fringe displaced from the central one by n cycles).

In very long-base-line measurements, the values of λ and D are usually of the order of 10 cm and 5000 km, respectively, so that the fringe spacing is typically of the order of 10^{-8} rad, or about 0.001 arcsec. if the motion of the source relative to the base line is due solely to the rotation of the earth, and if the base line is oriented strictly east-west, the fringes will be produced at the rate of about 2500 sec^{-1} . For every degree of motion, about half-a-million fringes will be produced. Such extremely high fringe density makes it very difficult to identify the central fringe, and this underscores one of the most serious problems encountered in attempting

* The fringe produced when the source direction is perpendicular to the base line.

to apply precision interferometry to high-accuracy determination of source position or base-line length. Assuming an uncertainty of ϵ_ϕ cycles in identifying the correct fringe, uncertainties ϵ_D and ϵ_θ in base-line length and source position, respectively, will be produced that satisfy the relation

$$\epsilon_\phi = \frac{\epsilon_D \cos \theta - D \epsilon_\theta \sin \theta}{\lambda} ;$$

and if $\theta \approx \pi/2$, then $\epsilon_\theta \approx -(\lambda/D) \cdot \epsilon_\phi$, which shows that the uncertainty in source position is of comparable order to the uncertainty in fringe number, expressed in radians (cf. Moran, 1968).

In practice, the fringe-position uncertainty is lessened by dividing the total record interval into a number of subintervals, on each of which the records are adjusted relative to one another so that the signals generated at both stations by the identical wave front are aligned. The adjustment can be quite closely calculated from estimates of the base-line geometry and from the known characteristics of the receiving equipment; corrections are applied later, as described below in the discussion of independent-clock interferometers.

The signals are received and recorded at as high a bandwidth as possible. Now, by virtue of the signal alignment, all frequencies in the bandwidth will be in relative phase agreement at, and only at, the time of alignment, at which time all harmonics will add. Thus, the first (or "white noise") fringe within each subinterval constitutes a global maximum of the amplitude and can be readily identified. When this condition prevails, and as long as $t \Delta\omega$ is small* compared to one cycle at the nominal frequency, the interferometer is described as operating in the white-fringe region.

* $\Delta\omega$ is the received signal bandwidth.

One of the principal disadvantages of the simple, or adding, interferometer is that it responds to the total power of large resolved features of the sky (Jennison, 1966). These do not produce fringes but may cause very large changes of the mean level of the signal. When the signal is amplified to obtain large fringes from a small, weak source, the erratic excursions of the recording system due to the background radiation may drive the fringe records off the recording scale. Ryle's multiplying, or phase-switched, interferometer is much more satisfactory from this point of view. In this arrangement, a half-wavelength section of cable is switched into and out of the feeder line from one antenna at a rapid frequency. This serves to alternate the phase of the signal from that antenna. The incoming signals from both antennas are added as in the previous case, so that when the cable section is omitted,

$$P_+(\theta) = |A(\theta)|^2 (1 + \cos \Delta\phi) ,$$

and when the cable section is inserted,

$$P_-(\theta) = |A(\theta)|^2 (1 - \cos \Delta\phi) .$$

Thus, a signal at the switch frequency is produced that is modulated by an envelope having the shape of the interference fringes. The factor $|A(\theta)|^2$, plus any slowly varying terms due to background noise, can be easily separated from the fringe terms by passing the signal through an amplifier tuned to the switch frequency, after which the switch frequency itself can be suppressed by a phase-sensitive detector stage.

What has essentially been accomplished is to subtract $P_-(\theta)$ from $P_+(\theta)$. The resulting response pattern,

$$P(\theta) = 2 |A(\theta)|^2 \cos \Delta\phi ,$$

is free from slowly varying background noise and can be much more highly stabilized.

Further, if the cable length is continuously varied at some fixed rate, the fringes, instead of alternating in position by $\pm \lambda/2D$ cycles, will steadily progress forward or backward at a speed that depends upon the rate at which the length of cable is varied. Actually, varying the feeder cable length in a channel amounts to changing the signal phase in that channel; in practice, it is usually much more convenient to change the phase by varying the local oscillator frequency instead. If the radio-frequency signal in the channel is nominally of the form $\cos(\omega_1 t + \Delta\phi)$ and if the local oscillator that beats down the signal is offset from a nominal frequency of ω_2 to one of $\omega_2 + \epsilon$, the product signal will be of the form $\cos[(\omega_1 - \omega_2)t + (\Delta\phi - \epsilon t)] + r.f.$ That is, the argument of the modified wave can be considered to have a phase that varies linearly with time.

This procedure is known as "rotating (or sweeping) the fringes," and the arrangement that utilizes this technique is generally referred to as a swept-frequency interferometer. The method finds important use in tracking source objects that move so fast through the fringes that they cannot be properly recorded. In such a case, the number of fringes can be decreased artificially so that enough of them can be seen to make meaningful measurements possible. This technique will also produce fringes when the base line is strictly north-south, even though no fringes would ordinarily be produced. Rotating the fringes with the source is desirable since this permits a compensation to be made for the high rate at which the source normally travels through the fringes. This procedure can be viewed as compensating for the doppler effects due to the differential rotation rates at the two stations.

The correlation interferometer employs still another technique for producing interference fringes. After amplification, the signals from both antennas are continuously multiplied in some sort of correlator. In a sense, this produces a response similar to that achieved in the phase-switched interferometer. Two signals, $e_1 = \cos \omega t$ and $e_2 = \cos (\omega t + \Delta\phi)$, multiplied together, produce

$$\epsilon = \cos \Delta\phi + \text{r.f.} \dots,$$

which is essentially the response noted above. In this case, however, there is no switching, and consequently no loss of useful signal. Since the product of two statistically independent noise currents has a zero mean, the correlator output will be unaffected by noise currents that are uncorrelated between the two channels, e.g., noise currents arising separately in the amplifiers. However, noise generated within the source under observation will be recorded by the interferometer. Like that of the switched-phase interferometer, the average response of this instrument to zero signal is itself zero.

3.2 Independent-Clock Interferometers

The independent-clock interferometer is a type of correlation interferometer that dispenses with cable or radio links between the two elements. Phase coherence between the two channels is maintained by providing each station with its own atomic clock. Hydrogen-maser and rubidium frequency standards have been employed, these being synchronized by use of Loran-C pulse transmissions, which can be employed for time calibrations accurate to $\pm 1 \mu\text{sec}$.

Signals are converted to video frequency, clipped, and sampled, so that a train of binary pulses is produced at each station. Each pulse train is recorded on digital tape, and both tapes are then correlated on a digital computer. Care must be taken to shift one set of pulses relative to the other by the proper amount to compensate for the differential doppler effects

described above. Any unpredicted delays can be compensated for by performing the correlations repeatedly for a number of values of delay, each of which differs slightly from the predicted value. The correlation function in each case is Fourier-analyzed, and the value of delay for which the spectrum peaks is taken to be the correct one.

Successful operation of independent-clock interferometers with very long base lines, typically measured in thousands of kilometers, has recently been achieved by several groups. Notable among these projects was the reported observation by Bare, Clark, Kellermann, Cohen, and Jauncey (1967) of four small-diameter radio sources on May 8-9, 1967, using an east-west base line of 4.61×10^5 wavelengths at 610 MHz between the 140-ft Tatel radio telescope at NRAO and the 85-ft antenna at the Maryland Point Observatory of the Naval Research Laboratory. Shortly thereafter, Moran, Crowther, Burke, Barrett, Rogers, Ball, Carter, and Bare (1967) reported a measurement conducted between the 120-ft Haystack radio telescope and the 140-ft NRAO telescope, forming a base line of 845 km, or 4.7×10^6 wavelengths at 1665.4 MHz. Other experiments have since been described in the literature: a four-element interferometer formed by antennas at Lincoln Laboratory, NRAO, University of California (at Hat Creek, California), and the Chalmers Institute of Technology (in Onsala, Sweden); another, using an interferometer base line of 1910 miles between the 150-ft telescope of the Algonquin Radio Observatory at Lake Traverse, Ontario, and the Dominion Radio Astrophysical Observatory (DRAO) 84-ft antenna near Penticton, B. C.; another, between the two NASA DSN stations near Woomera and near Canberra, Australia; and a recent link, between Canberra and the 210-ft dish at Goldstone, California.

Processing the received signals involves converting them to the video band after initial amplification at radio frequency. The video signal is itself amplified and then clipped, with the lower limit set at zero. The clipped wave form is sampled under control of some subharmonic of the natural frequency of the atomic oscillator. NRAO, for example, uses a bandwidth of 360 kHz, and samples at a 720-kHz rate, in accordance with the Nyquist theorem, while Lincoln Laboratory records at a bandwidth of either 6 or 120 kHz. Both groups record data continuously with time synchronization gaps inserted every 0.2 sec.

A correction must be applied to the correlator output in order to compensate for the error introduced by clipping the original wave form. If the raw output of the correlator is denoted by ρ_y , the corrected value ρ_x is given by

$$\rho_x = \sin\left(\frac{\pi}{2} \rho_y\right) .$$

This correction is due to van Vleck and is explained by considering $\rho_y \equiv \rho_y(\tau)$ to be the normalized statistical average of the product of the two pulse trains:

$$\rho_y(\tau) = \frac{\overline{y(t) y(t + \tau)}}{\overline{y^2(t)}} , \text{ where } y(t) = \pm 1 .$$

Now $\overline{y^2(t)} = 1$, since $y(t) = \pm 1$. Then the numerator term $\overline{y(t) y(t + \tau)}$ can be expressed as

$$\overline{y(t) y(t + \tau)} = \rho_y(\tau) = (+1) (P_{++} + P_{--}) + (-1) (P_{+-} + P_{-+}) ,$$

where P_{++} is the probability that both $y(t)$ and $y(t + \tau)$ are $+1$. The other terms can be similarly defined.

Turning to $x(t)$, the unclipped value, we can also define P_{++} by

$$P_{++} = \int_0^\infty \int_0^\infty p[x(t), x(t + \tau)] dx(t) dx(t + \tau) ,$$

where P_{++} is the joint probability that both $x(t)$ and $x(t + \tau)$ are positive. If $x(t)$ is assumed to have gaussian statistics, the function $p[x(t), x(t + \tau)]$ will have the bivariate gaussian distribution

$$p(x_t, x_{t+\tau}) = \frac{\exp \left\{ \frac{x_t^2 - 2\rho_x(\tau)x_tx_{t+\tau} + x_{t+\tau}^2}{2\pi R_x(0)[1 - \rho_x^2(\tau)]} \right\}}{2\pi R_x(0)[1 - \rho_x^2(\tau)]^{1/2}},$$

where $R_x(0) = \sigma_x \sigma_y$. Also, by symmetry,

$$P_{--} = P_{++},$$

$$P_{+-} = P_{-+}.$$

If $p(x_t, x_{t+\tau})$ is substituted into P_{++} , it can be integrated by transforming to cylindrical coordinates such that $x(t) = r \cos \theta$ and $x(t + \tau) = r \sin \theta$.

The result of this integration is

$$P_{++} = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho_x = P_{--}$$

It can also be shown that $P_{-+} = P_{+-} = (1/4) - (1/2\pi) \sin^{-1} \rho_x$, so that finally $\rho_y(\tau) = (2/\pi) \sin^{-1} \rho_x(\tau)$, which proves the result stated above.

Another correction is applied to compensate for the finite record length of each sample. Theoretically, the correlation function should be calculated as

$$R(\tau) = \int_{-\infty}^{\infty} x(t) \cdot x(t + \tau) dt.$$

However, in the practical situation, we have

$$R_{00}(\tau) = \int_{T-(\Delta t/2)}^{T+(\Delta t/2)} x\left(t - \frac{\Delta t}{2}\right) x\left(t + \frac{\Delta t}{2}\right) dt .$$

The correction is usually established by multiplying $R_{00}(\tau)$ by a weighting function $\omega(\tau)$, where $\omega(\tau)$ is chosen to satisfy the following criteria:

$$\omega(0) = 1 \quad (\text{normality}) ,$$

$$\omega(-\tau) = \omega(\tau) \quad (\text{symmetry}) ,$$

$$\omega(\tau)_{\tau \geq T_n} = 0 , \quad \text{where } T_n \text{ is the length of the record, i. e., the number of pulses contained within it.}$$

The last condition ensures that the corrected function $R_0(\tau)$ is defined for all τ . Even though $R_0(\tau)$ may not be a very good estimator of $R(\tau)$, its transform will generally be a satisfactory estimate of the smoothed value of the true spectral density of $R(\tau)$, provided the above three conditions are met. The function $\omega(\tau)$ is sometimes referred to as the "lag window," since it acts as a variable transmission factor whose weight varies with the delay parameter τ .

The object of applying this correction is to create a modified correlation function $R_0(\tau)$ whose transform exists and is equal in average value to the true transform. The function should be analytic in form, and it should vanish identically outside the range of τ within which the correlation function has meaning (i. e., $|\tau| \leq T_m$). It is desirable to weight the function more heavily near $\tau = 0$, where the calculated correlation function $R_{00}(\tau)$ can be expected to approximate most accurately the true function. Mathematically, the problem of choosing $\omega(\tau)$ so that the transform of its product with $R_{00}(\tau)$ enhances the region near $\tau = 0$ is analogous to the problem of designing an antenna whose response curve will concentrate the radiation near the central axis (Blackman and Tukey, 1958). To confine the window to a very narrow

band, however, will produce unwanted "side lobes" (in analogy with the antenna), and this will admit spurious signals. Therefore, $\omega(\tau)$ should be chosen smooth and gently varying to avoid this condition. Fortunately, in the case of the radio interferometer, the results are not very sensitive to the choice of lag function. For instance, many reasonable special cases of the function

$$\begin{aligned}\omega(\tau) &= a_0 + \sum_{j=1}^{\infty} a_j \cos \frac{j\pi\tau}{T_m} , & |\tau| < T_m , \\ &= 0 , & |\tau| > T_m ,\end{aligned}$$

can be found that are very acceptable. Two widely used functions are

$$\begin{aligned}\omega(\tau) &= \frac{1}{2} \left(1 + \cos \frac{\pi\tau}{T_m} \right) , & |\tau| < T_m , \\ &= 0 , & |\tau| > T_m ;\end{aligned}$$

and

$$\begin{aligned}\omega(\tau) &= 0.54 + 0.46 \cos \frac{\pi\tau}{T_m} , & |\tau| < T_m , \\ &= 0 , & |\tau| > T_m .\end{aligned}$$

The latter is known as the "Hamming" function, after R. W. Hamming of Bell Laboratories, who first proposed it. The former is sometimes called the "Hanning" function, after its inventor, J. von Hann, the Austrian meteorologist.

The object of the fringe-processing program is to derive the complex power spectrum, or visibility, of the interferometer as it responds to the radio source being tracked. Observing the long-term variation of the visibility function makes it possible to study the motion of the base line (which contributes to this variation). Although the foregoing discussion has, to simplify the exposition, concentrated on the analysis of monochromatic sources, any realistic power response will generally be of a spectral nature, i. e., $P = P(\omega)$. The power spectrum is a complex quantity that can be written as $P(\omega) = A(\omega)e^{j\Delta\phi(\omega)}$, where $\Delta\phi(\omega)$ is called the fringe amplitude, and $\Delta\phi(\omega)$ the fringe phase. Changes in base-line length contribute to the phase of the visibility, but not to its amplitude, and it is the former, therefore, that is of most interest in studies of base-line deformation.

Let us suppose that at some instant the source is equidistant from both antennas or, equivalently, that the signal received in one channel is delayed so that at that moment all frequencies within the bandwidth are in phase. If the ratio of the bandwidth to the nominal frequency is small, then all the frequencies will remain almost perfectly in phase during some subsequent time interval. Within this interval, the interferometer response will remain almost perfectly sinusoidal in shape, even though the bandwidth is of non-vanishing extent. When this circumstance prevails, the interferometer is said to be operating in the "white-fringe" condition; i. e., all the frequencies are present in the deep fringes.

It is convenient to consider that each frequency within the band generates its own fringe, and then to calculate as a figure of merit the time at which the fringe caused by the frequency furthest removed from nominal (viz., $\omega_1 = \omega_0 + \Delta\omega$) has gained one cycle relative to the nominal. This time interval serves to define, to an order of magnitude, the extent of the deep fringes. The figure of merit, expressed mathematically, is

$$n \frac{\lambda_1}{D} = (n + 1) \frac{\lambda_2}{D} \quad ,$$

which gives

$$n = \text{number of cycles} = \frac{\omega_1}{\Delta\omega} .$$

That is, the number of deep fringes is the ratio of the nominal radio frequency to the bandwidth (order of magnitude). A radio signal at 1 GHz recorded at a bandwidth of 100 kHz will yield 10^4 such fringes.

We have noted earlier, however, that a frequency error in one of the clocks relative to the other can cause the fringes to drift, so that the interferometer response assumes the form $\cos (\Delta\phi - \epsilon t)$. If this is indeed the case, as is to be expected during the actual measurements, then the fringe visibility will be reduced; i. e., the Fourier spectrum of the correlation function will be diminished in amplitude and "smeared," or leveled, across the frequency band (Figure 10).

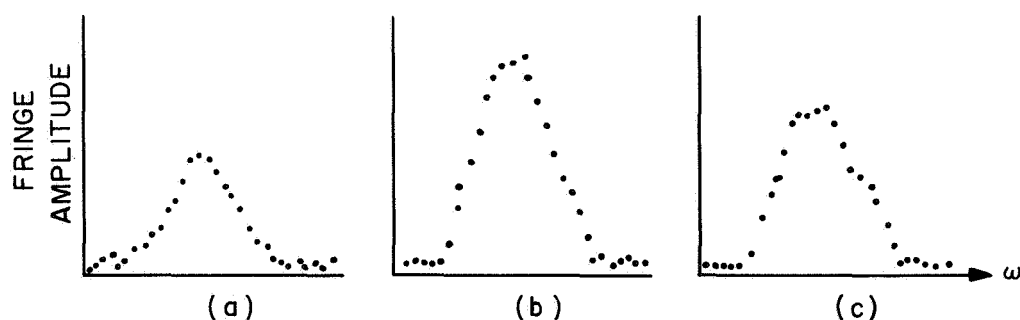


Figure 10. Sample fringe amplitude versus frequency for different relative frequency offsets: (a) 0.0 Hz; (b) 0.025 Hz; (c) 0.05 Hz. From Moran (1967).

It is usual to guess a number of trial delay offsets and to evaluate the Fourier spectrum of the fringes corresponding to each offset. The offset that causes the spectral response to peak (Figure 10b) is chosen to be the

correct value of delay, i. e., the value that establishes the white-fringe condition. To explain why this is so, let us consider the power spectrum $S(\omega)$ as it is obtained from the correlation function $R(\tau)$:

$$S(\omega) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^T R(\tau) e^{j\omega\tau} d\tau .$$

This will, in general, have the complex form

$$S(\omega) = A(\omega) e^{j\Delta\phi} e^{j\omega_T t} ,$$

where $\Delta\phi$ is the phase delay evaluated at the beginning of the recording interval, and ω_T represents the total fringe rotation rate during that interval due to all causes. The interval is chosen small enough that ω_T can be taken as a constant. Some effects that contribute to the presence of ω_T are the artificially induced fringe rotation, the rotational motion of the earth, and accumulated errors due to clock instability, etc. Terms such as the first two can be estimated readily enough; it is the last contribution that must be guessed. If a modified visibility function $S'(\omega)$ is defined as

$$S'(\omega) = \frac{1}{T} \int S(\omega) e^{j\omega'_T t} dt ,$$

where ω'_T is the trial function, which includes a guess for the unknown term, then it is clear that $S'(\omega)$ is maximized when the guessed ω'_T is equal to the true ω_T . For, as shown above, $S(\omega)$ is close to a sinusoid, and the integrated product of two sinusoids is nonvanishing when and only when their frequencies are equal. In practice (Moran, 1967), the offset can be determined by this method to better than 0.01 Hz, which amounts to an accuracy of 6 parts in 10^{12} in measuring the apparent difference between clock frequencies.

Usually the integrations with respect to τ and t are more conveniently reversed, for then the Fourier transform need not be evaluated so frequently.

Computation time involved in processing the fringes on a second-generation computer (Moran, 1968) currently runs at a ratio of about 1.5 hours per minute of record, using a 200-point correlation function, at 120-kHz bandwidth. About 75% of the computing time is occupied in performing the correlations, the rest in the transforms, bookkeeping, etc.

At this point it might be well to summarize briefly the main steps by which the determination of fringe phase is accomplished:

A. The radio-frequency signals are received, amplified, and converted to some convenient intermediate frequency (probably video). A wide bandwidth is desirable.

B. At each terminal the video signal is clipped so that all positive areas of its signal have a value of one, and all negative areas are zero. The clipped signal is then sampled at the Nyquist rate of twice the bandwidth, and the pulse trains that result are recorded on magnetic tape. With each record are included the recorder start time and the antenna pointing angles.

C. The tapes are brought together at a correlator, which may be a digital computer or possibly a computer of special design. For each record, the bit streams are shifted relative to one another by an amount that corresponds to the theoretically predicted geometrical delay, plus any fixed instrumental delay. This serves to produce white fringes. The records are then correlated. Usually it is advantageous to divide the correlation interval into a number of subintervals, in each of which the earth's rotation rate can be considered constant.

D. The correlation function for each subinterval is corrected for sampling and for finite record length. Then it is multiplied by the trial function $e^{j\omega_T \delta t}$, where ω_T is the sum of all frequencies that tend to rotate the fringes, e. g., earth's rate, fringe rotation rate, and other unknowns; and δt is the length of the subinterval. This is repeated for a number of guesses for the unknown phase rate.

E. Each of the modified correlation functions computed in Step D is subjected to Fourier transform. The case for which the Fourier spectrum peaks is taken to be the solution. For this case, the total phase delay that has been applied to the signal is recorded. From the long-term variation of this phase delay the deformation of the base line is to be inferred. The wider the bandwidth, the fewer are the deep fringes, and therefore the more precisely the white-fringe region can be identified.

The clocks can be synchronized to very high accuracy by using pulses received from Loran-C navigation transmissions. Basically, the Loran-C system comprises a number of station chains, located throughout the world, that emit coded signals that are very accurately phased in time. For example, a ship receiving signals from any two sets of stations can determine its own position as lying on the intersection of two hyperbolas. The parameters of each hyperbola are determined by finding the elapsed time between the receipt of corresponding pulses from both stations of a pair. Or, expressed alternatively, a hyperbola is the locus of all points for which the difference in distance (or time, here) from a given point is constant.

The transmissions use a 100-kHz carrier frequency, which permits high accuracy owing to low ground-wave attenuation. Pulse lengths are about 300 μ sec at a bandwidth of 20 kHz. All stations code their pulse trains differently, so that each can be identified. Measurements are made in three steps. First, the ambiguity as to which pulse is being observed is resolved by reference to some relatively coarse external timing signal, such as standard HF or VLF time transmissions. Next, the station clock is set to the Loran-C transmissions by comparing the arrival of any particular Loran-C pulse with a time of coincidence (TOC). TOC's are the times at which the first pulse of a transmission from a Loran-C master station coincides with the UTC second as calculated at the U. S. Naval Observatory (USNO). (For the East-coast Loran-C chain, TOC does indeed occur once each second.*) The final time adjustment is made by identifying individual

*Propagation-delay estimates are available from the USNO, on request.

cycles within a pulse. If proper cycle selection is accomplished visually (Shapiro, 1968), synchronization can be achieved to within about 1 μ sec. Phase-tracking receivers can then automatically identify the cycles within the pulse to as close as 0.1 μ sec.

An alternative method for maintaining clock synchronism was suggested by Gold (1967), who proposed using the interferometer itself to regulate the oscillators. For simplicity, let us assume that a narrow-diameter radio source can be located at precisely 90° declination. Then, whenever an east-west interferometer views it, the angle between the source vector and the base line would be precisely $\pi/2$. Over an indefinite period, therefore, the fringes should remain stable unless the oscillators drift relative to each other. It should then be possible to stabilize the clocks by periodically (say, once an hour) viewing the source and regulating the clocks to hold the fringes stationary. At source frequencies of 1 GHz, it should be possible to synchronize the two clocks to better than 1 nsec, provided the fringes can be stabilized to at least one fringe width.

As was noted earlier, the total fringe phase is expressible in terms of the base-line geometry, certain instrumental delays, and various small effects that cannot be quantitatively predicted. The total phase can be expressed as

$$\Delta\phi = \frac{D}{\lambda} \cos \theta + \eta = \frac{D}{\lambda} (\vec{B} \cdot \hat{S}) + \eta ,$$

where B is the base-line vector, S is the unit vector in the direction of the source, and η is the sum of all other effects that contribute to the total phase.

Let us suppose that B has angular coordinates (L, ℓ) relative to the equator, where $L = \omega_e t + L_0$, with ω_e being the earth's diurnal rotation rate and L_0 measured from the vernal equinox. If the source direction \hat{S} is characterized by the angular coordinates (α, δ) , also measured relative to an equatorial reference system, then

$$\vec{B} \equiv \frac{D}{\lambda} (\cos \ell \cos L, \cos \ell \sin L, \sin \ell) ,$$

$$\hat{S} \equiv (\cos \delta \cos \alpha, \cos \delta \sin \alpha, \sin \delta) ,$$

and thus, according to the relation above,

$$\Delta\phi = A + B \cos \omega_e t + C \sin \omega_e t ,$$

where

$$A = \frac{D}{\lambda} \sin \delta \sin \ell + \eta ,$$

$$B = \frac{D}{\lambda} \cos \delta \cos \ell \cos (\alpha - L_0) ,$$

$$C = \frac{D}{\lambda} \cos \delta \cos \ell \sin (\alpha - L_0) .$$

Values for A , B , and C can be estimated by fitting a sinusoid to a number of observations of $\Delta\phi$. The last three equations relate the five unknowns D , δ , ℓ ($\alpha - L_0$), and η to one another. If two more sources are simultaneously observed from the same base line, a total of nine equations exists, relating nine unknowns (the above five, plus $\alpha - L_0$ and δ for each of the two other sources) to the nine measurables A_i , B_i , and C_i , for $i = 1, 2, 3$ (Moran, 1968). For present purposes, it would probably be more convenient to express $\Delta\phi$ in perturbational form, viewing its change as a function of small changes in the five unknowns,

$$d(\Delta\phi) = \frac{\partial(\Delta\phi)}{\partial D} dD + \frac{\partial(\Delta\phi)}{\partial \delta} d(\delta) + \dots ,$$

in which case the coefficients of the sinusoid in $\omega_e t$ would become functions of the perturbed quantities dD , $d\delta$, \dots .

The fringe rate is inversely proportional to the fringe period (in the θ -domain, as well as in the t -domain), which was shown earlier to be equal to $\cos^{-1}(\lambda/D)$ rad. This suggests that values of the base-line length alone might be obtained directly from the measurement of fringe frequency, without the necessity for solving the complete set of simultaneous equations containing all the unknowns. However, the direction of base-line deformation cannot be found this way, which substantially limits the usefulness of this particular approach.

3.3 Use of an Artificial Source

It is natural to consider the possibility of mounting a transmitter aboard a satellite, or perhaps even on the moon, to act as a radio source for the interferometer.* By using available electronics and power technology, we might expect to experience stronger signal reception over such relatively small ranges; appropriate orbit choices should provide frequent and favorable periods of visibility from ground stations advantageously placed with respect to the crustal motions anticipated in specific regions; and, with the employment of suitably coded signals of broad bandwidth, it should be possible to reduce the fringe ambiguities substantially.

What is perhaps the most serious drawback to the concept is the uncertainty in source position. For the satellite, this results from imperfect knowledge of the perturbations on its orbit by the geoid, radiation forces, outgassing, and lunisolar forces, to mention a few sources of error.† For the moon, the uncertainty is due to the imprecision of the lunar ephemeris and to the lunar librational motion, which can produce large positional displacements of the source.

* Recently, a long-base-line experiment was conducted between Canberra and Goldstone, during which Pioneer 8 served as the source (D. S. Robertson, 1968, private communication to C. A. Lundquist).

† However, this suggests the possibility of using the interferometer method for accurately tracking satellites, if the base line is known well enough.

Since the source is not at an infinitely great distance from the earth, the simple relation $\Delta\phi = (D/\lambda) \cos \theta$ no longer holds, and a corrected equation must be supplied. Referring to Figure 11a, we see that

$$R \cos \gamma = R + d - D \cos \theta \quad ,$$

from which

$$d = D \cos \theta - D \sin \theta \left(\frac{1 - \cos \gamma}{\sin \gamma} \right) \quad ,$$

where γ is the angle subtended at the source by the two range vectors. Thus,

$$\Delta\phi = \frac{D}{\lambda} \left[\cos \theta - \sin \theta \left(\frac{1 - \cos \gamma}{\sin \gamma} \right) \right]$$

represents the corrected equation for $\Delta\phi$.

If the vector from the first station to the source is denoted by \vec{R}_1 and that from the second station to the source by \vec{R}_2 , and if vectors from the center of the earth to the two stations and to the source are given by \vec{S}_1 , \vec{S}_2 , and \vec{S} , respectively, then (Figure 11b)

$$\vec{R}_1 = \vec{S} - \vec{S}_1 \quad ,$$

$$\vec{R}_2 = \vec{S} - \vec{S}_2 \quad ,$$

so that

$$R_1 R_2 \cos \gamma = \vec{R}_1 \cdot \vec{R}_2 = |\vec{S}|^2 - \vec{S} \cdot (\vec{S}_1 + \vec{S}_2) + \vec{S}_1 \cdot \vec{S}_2 \quad .$$

This last expression yields γ in terms of the station and source positions, from which it is possible to determine $\Delta\phi$ by the equation cited above.

The elevation angle e at any station can be expressed in terms of source range and angular separation between the subsource point and the station (cf. Figure 11c) as

$$R \cos (\psi + e) = R_e \cos e \quad .$$

It is usually the case with radio or optical tracking that atmospheric limitations and terrain features dictate some minimum value of e that can be tolerated. For laser ranging, values of e less than about 30° cannot be considered. The above equation, then, gives a side condition that relates R_i and ψ_i as they appear in the expression for $\cos \gamma$ (and hence in the expression for $\Delta\phi$). Explicitly,

$$\cot \psi - \frac{R_e}{R \sin \psi} = \tan e \quad .$$

This equation can also be used to derive the time-in-view relation, given the two sets of station coordinates and the source ephemeris. In the present analysis, the source must be in simultaneous view of both stations for the observation to be meaningful.

The stellar interferometer furnishes information on base-line length and direction only, and not on the absolute position of any of its points. Thus, for example, if a series of stellar measurements were to show no change in the base-line vector, it would be impossible to determine (without other information) whether the base line had truly suffered no change in actual position, or whether it had been translated parallel to itself. This question is important for geophysical measurements, not because of any expectation that base-line vectors will show strictly parallel displacements, but rather because base-line deformation will generally be due to a number of forces, such as those of tidal origin, earth tremors, or (perhaps) true continental drift. Knowing only the vectorial change of the base line does not suffice to separate radial from transverse components of the motion.

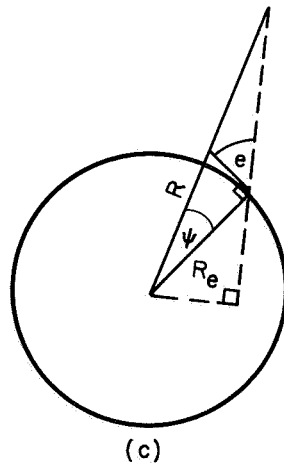
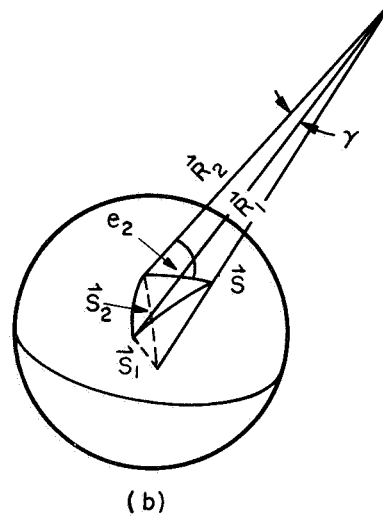
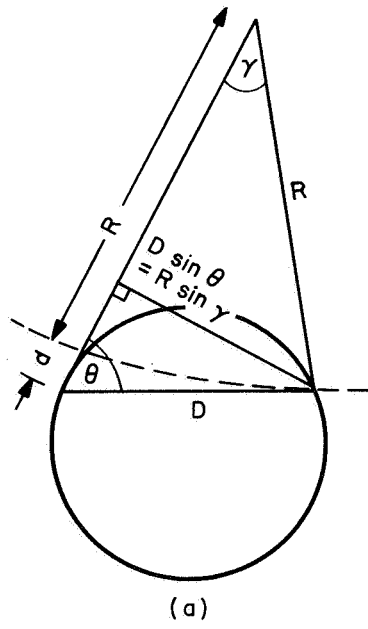


Figure 11. Base-line — satellite geometries for various aspects of satellite interferometry.

Satellite or lunar interferometry, on the other hand, brings the actual position of the base-line vector into play through the variable γ as it appears in the term $(D/\lambda) \sin \theta [(1 - \cos \gamma)/\sin \gamma]$ found in the fringe-phase equation above, where γ is related to R by (cf. Figure 11a)

$$R \sin \gamma = D \sin \theta ,$$

and R depends upon the source's (and the stations') absolute distances from some reference point. So it becomes evident that in order to make use of this type of interferometer, it will be necessary to measure some characteristic length to sufficient precision. This length might be the source's distance from the earth's center of mass, or the source-antenna range. Laser ranging holds promise of permitting distance measurements of sufficient accuracy to be made. With ruby lasers, pulse lengths less than 10 nsec have been achieved in the laboratory; reduction to the order of picoseconds appears feasible. Even with nanosecond pulse widths, the laser can be expected to yield round-trip transit times to a reflecting satellite that are also accurate to the order of a nanosecond, which is equivalent to a 15-cm uncertainty in the one-way range (Bender, Alley, Currie, and Faller, 1968).

4. LASER SYSTEMS

C. G. Lehr

4.1 Present Laser Systems

4.1.1 Introduction

Satellite orbits with meter accuracy could open up an investigation of crustal motions, and, as discussed above, orbits with decimeter accuracy would make the detection of continental drift an attractive possibility.

However, these accuracies cannot be obtained by using the Baker-Nunn camera for satellite tracking. The reason is that the 2- to 3-arcsec observational error of the cameras limits the results of the orbit computation to an accuracy of about 10 m. The situation is different when laser systems are used to track satellites. Like radars, they measure the distance to the satellite rather than its direction in space, and at present they measure these distances to within 1 or 2 m. Future accuracies of decimeters may be possible.

Laser systems have become practical in the last few years because six satellites (two French and four American) now in orbit were equipped with retroreflectors. Each retroreflector is made up of an array of hundreds of corner cubes that concentrate the returning laser radiation sufficiently to make laser ranging feasible. In fact, the equipment needed on the ground and on the satellite turns out to be relatively simple and inexpensive when compared with other electronic tracking systems. Of these six satellites, only Geos 1 and Geos 2 can be ranged on in both the Northern and Southern Hemispheres. These two satellites are gravitationally stabilized; consequently, their retroreflectors point toward the earth wherever they are in their orbit. The other four satellites are stabilized by the earth's magnetic field and their retroreflectors point toward the earth only when they are over the Northern Hemisphere.

Since lasers operate within the visual spectrum, they, like the satellite-tracking cameras, require cloudless skies. They should therefore be located at good astronomical observing sites. In contrast to the cameras, however, they can be used when the satellites are in the earth's shadow and when the sky is not dark. Although increased accuracy is the main reason for using laser systems for satellite tracking, there is a secondary advantage: The data can be obtained quickly. The time of flight of the pulse of laser energy from the laser to the satellite and back again is presented immediately in digital form by a time-interval counter. The atmospheric correction is small in magnitude and simple in form. This rapid availability of the data should lead to improved predictions of a satellite's angular coordinates and, in turn, to an increase in the system's maximum range and to an improvement in its accuracy.

4.1.2 Retroreflectors on satellites

Three different arrays of retroreflectors are in current use: the ones on BE-B and BE-C, on D1-C and D1-D, and on Geos 1 and Geos 2. Each was designed to spread the returning beam enough to compensate for the angular deflection due to velocity aberration. This spread is introduced because the axis of the returning beam intersects the earth at a distance up to about 100 m from the laser system; the point of intersection varies from one satellite pass to another. If the beam were not spread, the photoelectric receiver would have to be separated from the laser transmitter, and the receiver would have to be moved to a new position for each satellite pass. The characteristics of the six satellites and their retroreflectors are given in Table 6.

4.1.3 Principles of operation

Laser tracking systems are optical radars. A pulse is transmitted to the satellite; it is reflected and the return is detected. The data consist of the epoch at which the laser is fired and the time taken for the pulse to travel to the satellite and return. A correction for the reduced velocity of

Table 6. Satellites with retroreflectors

Satellite	Apogee (Mm)	Perigee (Mm)	Inclination (deg)	Period (min)	Average visual magnitude	Number of corner cubes	Effective area (cm ²)	Divergence (arcsec)	Stabilization
BE-B (6406401)	1.09	0.89	80	105	8	360	80	12	magnetic
BE-C (6503201)	1.32	0.94	41	108	8	360	80	12	magnetic
Geos 1 (6508901)	2.27	1.12	59	120	8	334	940	20	gravitational
D1-C (6701101)	1.35	0.53	40	104	8	144	20-100*	16*	magnetic
D1-D (6701401)	1.85	0.58	39	110	8	144	20-100*	16*	magnetic
Geos 2 (6800201)	1.61	1.08	74	112	8	400	1100	20	gravitational

*The retroreflectors on D1-C or D1-D shape the returning beam so as to compensate most effectively for velocity aberration. Consequently, a characterization of its properties in terms of effective area and divergence is an approximation.

light in the atmosphere is made. The magnitude of this correction is 2 to 6 m. If the ambient temperature and pressure are measured at the time of the observation, the correction can be determined to about 1% (Lehr, 1968a). Still more accuracy in the atmospheric correction can be obtained by making two simultaneous laser measurements at different wavelengths (Owens and Bender, 1967).

When time is converted to range, the value of c , the velocity of light in vacuum, is used. Since the experimental accuracy of c is, at present, only about 1 part in 10^6 , a sizable range error is introduced. However, this error turns out to be unimportant when, as in the present application, small changes in range among successive measurements, rather than a single accurate measurement, are of interest. For proper correspondence with the measured range, the true epoch is the instant at which the laser pulse reaches the satellite. It is obtained by adding half the travel time of the laser pulse to the measured epoch. Usually this correction is not made until the orbits are computed. For satellite tracking, optical radar has the following advantages over microwave radar:

- A. The inherently short time duration of the pulse favors high accuracy.
- B. The narrow beamwidth improves the efficiency.
- C. The short wavelength simplifies the atmospheric correction.

The disadvantages with respect to microwave radar are the difficulty of operating in daylight and the impossibility of operating in cloudy weather, but radar corrections for clouds degrade the accuracy of their measurement.

4.1.4 A comparison of laser systems

The characteristics of some geodetic laser-ranging systems are listed in Table 7. The precision of the range measurement can be estimated by use of the fact that a two-way travel time of 1 nsec corresponds to 15 cm. The system's range is proportional to the one-fourth power of the transmitted energy, inversely proportional to the square root of the transmitter beamwidth, and directly proportional to the square root of the diameter of the

Table 7. Characteristics of geodetic laser-ranging systems

System	GSFC (Maryland)	CNES (France and Spain)	AFCRL (Massachusetts)	Hawaii	Greece	Arizona
Pulse duration (nsec)	12-18	25-30	30	40-60	10	15
Pulse energy (J)	1	1	1.8	0.5	1	7.5
Power output (Mw)	70	30	60	8	100	500
Pulse-repetition rate (min^{-1})	60	15	20*	2	2	1
Counter resolution (nsec)	1	10	1	10	10	1
Transmitter beamwidth (mrad)	1	0.5-2	3	1	1	0.6-6
Receiver aperture (inches)	16	14	5	16	16	20
Receiver bandwidth (nm)	1	1.8	1	2	2	0.6
Receiver beamwidth (mrad)	1.5-5	0.5-2	1-5	2	2	0.6-6
Tracking	programmed	visual	programmed	visual	visual	positioned

*Each transmission is a burst of 10 pulses, 60 nsec apart.

GSFC: Goddard Space Flight Center, National Aeronautics and Space Administration, U. S. A.

CNES: Centre National d'Etudes Spatiales, France.

AFCRL: Air Force Cambridge Research Laboratory, U. S. A.

SAO: Smithsonian Astrophysical Observatory, U. S. A.

receiver aperture. The sky-background current is directly proportional to the receiver's optical bandwidth and to the square of its beamwidth. With programmed tracking, the system continuously tracks satellites at night or in daylight, whether they are within or outside the earth's shadow. Visual tracking can be used only at night and on sunlit satellites. Positioned tracking can be used under the same conditions as programmed tracking, but the pulse-repetition rate must be low because the tracking mount is moved manually to a discrete series of angular positions. The Air Force Cambridge Research Laboratory (AFCRL) laser transmits a burst of pulses. This mode of transmission compensates for an observed scintillation, or amplitude variation, in the returns (Morrison and Ackerman, 1967).

4.1.5 The laser system on Mt. Hopkins

The laser system on Mt. Hopkins was designed to be a prototype of one that could be installed at each of the 12 Smithsonian astrophysical observing stations. Its features reflect the particular capabilities and needs of the Smithsonian program of precise orbit computation. Its characteristics and operation can be used to illustrate topics discussed in subsequent sections.

The system has a static pointing mount rather than a tracking mount. It uses predictions of a satellite's azimuth and elevation instead of relying on visual tracking. The accuracy of the predictions is several arcmin, a value that should be well within the maximum 6-mrad (20-arcmin) beamwidths of the laser and the receiving telescope. The static pointing mount was selected because it permits the laser system to operate when the satellite is in the earth's shadow and it may permit operation during daylight hours. Because the mount is manually operated, the pulse-repetition rate is low. But this rate should be adequate for the computation of precise orbits when the network of stations is complete. The operational simplicity of the system may be desirable in some of the network's more remote locations.

4.1.6 Discussion

We have referred above to the fact that the reduction of the laser range observations can be accomplished easily and rapidly. We now explain how this reduction is done. Two corrections are needed. The first corrects for time delays in the system that make the reading of the range counter larger or smaller than that corresponding to the difference between the intersection of the mount's azimuth and elevation axes and the retroreflector on the satellite. The second corrects the velocity of the laser pulse, which is less in the earth's atmosphere than it is in vacuum.

At Mt. Hopkins, the correction for the time delay is obtained by ranging on a target 2547 ft away from the laser system. The distance, determined from an accurate survey, is subtracted from the product of the counter reading and the atmospheric velocity of light. The difference is added to the measured range values.

The correction for the earth's atmosphere comes from the work of G. D. Thayer (1967, private communication). It is based on a model that uses the height of the observing station, the local temperature and pressure, and the elevation of the satellite.

If r_v is the travel time of the laser pulse multiplied by half the velocity of light in vacuum and r_m is the "measured range" (i. e., r_v corrected for the effect of the atmosphere), we have

$$r_m = r_v - \frac{2.238 + 0.0414PT^{-1} - 0.238h}{\sin \alpha + 10^{-3} \cot \alpha} s, \text{ for } \theta_0 > 5^\circ, \lambda = 694.3 \text{ nm},$$

where

r_m and r_v are range values in meters,
P is the atmospheric pressure in millibars,
T is the temperature in degrees Kelvin,

h_s is the laser's elevation above mean sea level in kilometers,
 α is the elevation angle of the satellite, and
 θ_0 is the apparent elevation angle (i. e., the elevation angle uncorrected for atmosphere bending).

The standard error associated with this correction is 5 cm or less for elevations greater than 20° .

The resolution of the Mt. Hopkins system, for example, lies between the half-pulse duration, 8 nsec (1.1 m), and the counter resolution, 1 nsec (0.15 m). The higher value is obtained when the return is weak and not enough photoelectrons are generated to define the leading edge of the received pulse. The lower value is obtained when the return is strong, large numbers of photoelectrons are produced, and the shape of the transmitted pulse is faithfully reproduced. Then one can determine where on the pulse's leading edge the time-interval counter is triggered. Statistical factors are involved in considerations of the strength of the return signal. Its intensity varies randomly, even when the range is fixed. The variation may occur because the retroreflector is multifaceted and the radiation is coherent (Goodman, 1965; Lehr, 1968b). The intensity of the returning radiation varies randomly over the cross section of the beam. Since a correlation cell is much larger than the aperture of the receiving telescope, there is no aperture averaging and the statistical variation of the received signal strength is large.

4.2 Possible Improvements in Laser Systems

4.2.1 The effect of the detection scheme

We stated above that the resolution of a laser system is expected to be greater when the returned signal is stronger. The reason is that in this case the resolution of the time-interval counter is greater than the pulse duration of the Q-switched laser. As the strength of the signal increases, the quantization of the return, which is due to the discreteness of the

electronic charge, introduces less distortion into the shape of the pulse. Consequently, corresponding points on the transmitted and received pulses can be more precisely identified with each other.

We now consider the effect of signal strength in a quantitative manner. Two cases are important. They represent different ways of identifying a particular point within the pulse. The first method uses a time-interval counter as described above. The counter is triggered when the leading edge of the received pulse reaches a given threshold. The particular instant at which this triggering occurs depends on the amplitude, or total energy, of the received signal, but the point is well defined if the energy or amplitude of each return is measured. The second method uses a high-frequency oscilloscope to display the return pulse. Currently, this method is used to confirm and analyze returns, but not as yet to obtain precise values of the transit time. However, if the start of the sweep were precisely timed and the sweep rate were accurately known, it could be used for this purpose. The procedure is the following: The presentation on the screen of the oscilloscope is photographed, and the centroid of the pulse is determined by considering the return to be a histogram of the photoelectrons. In this way, the statistical irregularities in the pulse are dealt with most effectively. An important advantage of this method over threshold detection is the fact that the reference point on the pulse is independent of the strength of the signal. It may turn out that electronic methods could be used to determine the centroid of the pulse. But the method of oscilloscope display would be at least as accurate as any other, although it would probably be slower.

In a quantitative consideration of threshold detection, we assume that the counter is triggered at the point of maximum slope on the pulse's leading edge. Here the resolution is best. The resolution will be poorer when the relative values of signal energy and threshold level cause the counter to be triggered elsewhere. We also assume that the transmitted pulse has the shape of a normal (gaussian) density function. This assumption is a good one for a Q-switched laser. If the system introduces no distortion, the shape of the received pulse tends to that of the transmitted pulse as the

number of detected photoelectrons increases. We confine our analysis to a distortionless system.

We express the normal density function in the form that is used in the theory of probability:

$$\phi(x) = (2\pi)^{-1/2} \exp\left(\frac{-x^2}{2}\right) , \quad (1)$$

where the mean value of x is zero and its standard deviation is one. If $\phi(x)$ is in watts, equation (1) represents a pulse whose energy is

$$E = \int_{-\infty}^{\infty} \phi(x) dx = 1 \text{ J} . \quad (2)$$

The variable x is proportional to time. The maximum value of equation (1), occurring when $x = 0$, is

$$\phi_{\max}(x) = \phi(0) = (2\pi)^{-1/2} = 0.400 . \quad (3)$$

For the half-power points we have

$$0.2 = (2\pi)^{-1/2} \exp\left(\frac{-x^2}{2}\right) , \quad (4)$$

with the solution $x = \pm 1.18$, which corresponds to a pulse width of 2.36. Consequently, the pulse width is 2.36 times the standard deviation.

The first and second derivatives of equation (1) are

$$\phi'(x) = -(2\pi)^{-1/2} x \exp\left(\frac{-x^2}{2}\right) , \quad (5)$$

$$\phi''(x) = (2\pi)^{-1/2} (x^2 - 1) \exp\left(\frac{-x^2}{2}\right) \quad (6)$$

The slope has its maximum value when $\phi''(x) = 0$, or $x = -1$. The maximum value of the slope is

$$s = \phi'(-1) = (2\pi)^{-1/2} \exp\left(\frac{-1}{2}\right) = 0.242 \quad . \quad (7)$$

At the point of maximum slope, the amplitude of the return is $\phi(-1) = 0.242$.

We continue our consideration of the case where the threshold level is set at the point of maximum slope. If the signal is a single photoelectron, it can be emitted anywhere within the duration of the pulse, with highest probability near the pulse's midpoint and with lowest probability at the extreme edges. In this situation, the shape of the returned pulse bears little resemblance to the transmitted pulse. The error (standard deviation) of the measurements is $\sigma_x = 1$ or $\sigma_t = \tau/2.36$, where τ is the pulse duration. The corresponding range error is $\sigma_r = 15\sigma_t = 6.35$ cm, when τ is expressed in nanoseconds. When the signal consists of more electrons, the definition of the pulse shape is better, and the point where it activates the time-interval counter is nearer to the threshold setting of maximum slope. For a quantitative measure of improved accuracy, we are interested in the case where the counter is triggered within an interval of duration $h\tau$, where $h < 1$ and smaller values of h indicate more precise range measurements. We let n be the mean value of the number of photoelectrons generated within the time $h\tau$. The actual number of electrons within this time will differ from n . With use of Poisson statistics, the standard deviation of n , or square root of its variance (Feller, 1968, p. 228), is $n^{1/2}$. For large n , the Poisson law is well approximated by the normal law, for which tables show that with a 95% probability the signal will be within $\pm 2n^{1/2}$ of n . Or $4n^{1/2}$ electrons must be received for the probability to be 95% that the return time is known to within $h\tau$.

Using Figure 12 and the facts that the maximum slope is 0.242 and that the maximum slope occurs at $x = 0.242$, we obtain the following equation:

$$4n^{1/2} = 0.242(2.36h) \left(\frac{n}{0.242} \right) \quad , \quad (8)$$

where the factor $n/0.242$ changes the scale of the $\phi(x)$ curve to give n electrons within a time $h\tau$ centered at the point of maximum slope. From equation (8) we obtain

$$n = \frac{2.88}{h^2}, \quad \text{for } h \ll 1 \quad (9)$$

This equation gives the number of electrons that must be received within $h\tau$ if there is to be a 95% probability that the precision of the range measurement is $h\tau$. For example, let us consider $h = 0.1$: The time is measured to 0.1 the duration of the laser pulse. For a 15-nsec pulse, the precision would be 1.5 nsec (or 22.5 cm). Then, according to equation (9), $n = 288$ photoelectrons.

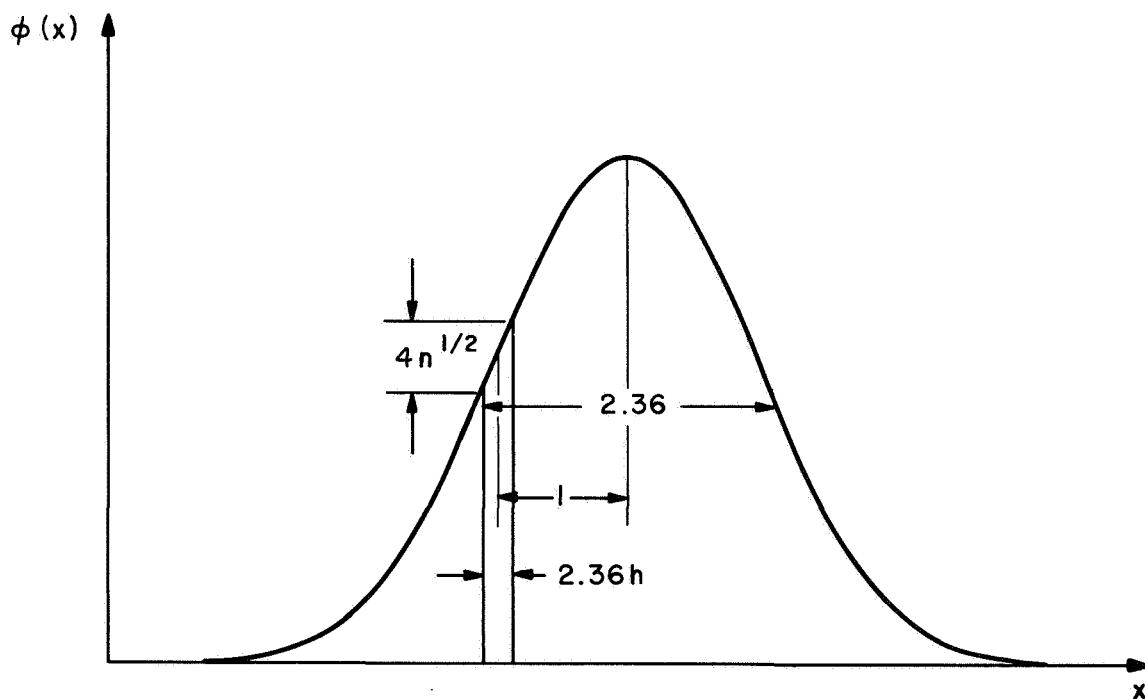


Figure 12. Diagram for threshold detection near the point of maximum slope.

To find N , the total number of electrons in the pulse, we divide n by the ratio of the area of the strip in which the n electrons are generated to the total area under the $\phi(x)$ curve; that is,

$$\frac{n}{N} = \frac{\int_{1-1.18h}^{1+1.18h} \phi(x) dx}{\int_{-\infty}^{\infty} \phi(x) dx} = \Phi(1+1.18h) - \Phi(1-1.18h) \quad , \quad (10)$$

where

$$\Phi(x) = \int_{-\infty}^x \phi(u) du \quad \text{and} \quad \Phi(\infty) = 1 \quad ;$$

or, using the approximation that the upper bound of the strip is a straight line whose slope is 0.242 and the fact that the width of the strip is 2.36h, we obtain

$$\frac{n}{N} \doteq (0.242)2.36h = 0.570h \quad , \quad \text{for } h \ll 1 \quad . \quad (11)$$

For $h = 0.1$, $n/N = 0.0570$ and $N = 5050$ electrons.

We now consider centroid detection. We assume that the return has been displayed on an accurate time base and that individual photoelectrons are resolved. The display is considered to be a histogram of the electrons. The instant of pulse reception is set equal to the average value (centroid) determined from this histogram. The travel time of the laser pulse is the time between the centroids of the transmitted and received pulses.

Again we ask how many photoelectrons must be received for 95% probability that the epoch of the returned pulse is within $h\tau$. We use the fact that the electrons are generated independently and, consequently, the variances in their arrival times are additive. The standard deviation σ for a single electron is simply that associated with the transmitted pulse, which has the shape of a normal density function. The value of σ is $\tau/2.36$. The

standard deviation of the average of N independent electrons is, from the fact that their variances are additive,

$$\sigma_N = \frac{\sigma}{N^{1/2}} = \frac{\tau}{2.36 N^{1/2}} \quad (12)$$

Since the average value also has a normal distribution, the probability is 95% that it is within $2\sigma_N$ of the true midpoint of the pulse. Or, with 95% probability, the condition for the precision of the return to be within $h\tau$ is

$$4\sigma_N = h\tau \quad \text{or} \quad \frac{4\tau}{2.26 N^{1/2}} = h\tau \quad (13)$$

The number of received electrons required for a given h must then be

$$N = \frac{2.88}{h^2} \quad (14)$$

For comparison with the previous method, we let $h = 0.1$, a value that gives a range error of 22.5 cm for a 15-nsec pulse. The required N is 288 electrons. For a still smaller range error of 1 cm, h must be $0.1/22.5 = 4.45 \times 10^{-3}$, and N is then 146,000 electrons.

The signal strengths required for centroid detection come out to be much lower than those required for threshold detection. This result is consistent with the fact that centroid detection is an idealized procedure that distinguishes every one of the returning electrons. Threshold detection, on the other hand, is a practical technique already in use.

Let us first use the formula we have obtained for threshold detection to examine the limiting accuracy of the laser system at Mt. Hopkins and then go on to considerations involved in a system of greater precision.

The precision of the laser system at Mt. Hopkins is limited by the 1-nsec resolution of the range counter. We assume that the counter reads to the nearest nanosecond, so the minimum range error for a single return corresponds to 0.5 nsec, or 7.5 cm. Since the pulse duration is 15 nsec, $h = 0.5/15 = 1/30$. From equations (9) and (11) we see that the required signal strength for threshold detection is

$$N = \frac{2.88/h^2}{0.570h} = \frac{5.05}{h^3} \text{ electrons} , \quad (15)$$

or $N = 5.05 \times 30^3 = 1.30 \times 10^4$ electrons.

At what satellite range can we obtain this signal with the laser system at Mt. Hopkins? We use the range equation

$$S = E \cdot \frac{A_s A_R}{\Omega_T \Omega_S} \cdot T^2 \cdot \frac{10^{-5}}{2.86} \cdot \frac{1}{R^4} , \quad (16)$$

where

- S is the number of photons arriving at the aperture of the telescope,
- E is the transmitted energy (7.5 J),
- T is the atmospheric transmission (0.7),
- A_S is the effective reflecting area of the satellite ($9.35 \times 10^{-2} \text{ m}^2$ for Geos 1),
- Ω_S is the solid angle of the beam reflected from the satellite ($7.3 \times 10^{-9} \text{ sr}$ for Geos 1),
- A_R is the area of the telescope's aperture (0.202 m^2),
- Ω_T is the solid angle of the transmitted laser beam ($2.66 \times 10^{-7} \text{ sr}$, corresponding to 2 arcmin), and
- R is the satellite range in megameters.

Inserting numerical values into equation (16), we have

$$R = \frac{105.7}{s^{1/4}} \quad (17)$$

The value of S is obtained from N by making these assumptions:

- A. The excess signal loss^{*} is 20 dB.
- B. The quantum efficiency of the photomultiplier is 3%.
- C. The optical efficiency of the telescope is 40%.

The use of these assumptions gives

$$S = \frac{N}{0.01 \times 0.03 \times 0.40} = 8.33N \times 10^3, \\ = 8.33 \times (1.36 \times 10^4) = 1.132 \times 10^9 \text{ photons,}$$

and $R = 1.057/11.32^{1/4} = 0.573 \text{ Mm}$, which is just about the perigee of D1-C, or the lowest range for any retroreflecting satellite now in orbit. Thus, the counter resolution does not limit the precision of the present system.

Instead, its precision is a function of the range of the satellite. If $N = 3$, we have a 95% probability that one or more electrons is generated on any return. This value corresponds to $S = 2.5 \times 10^5$ photons or a range of 4.7 Mm. At this range, there is 95% probability that a single range measurement will be within

$$2\sigma_t = \frac{\tau}{1.18} = \frac{15}{1.18}, \\ = 12.7 \text{ nsec}$$

We then have $2\sigma_r = 12.7\tau = 12.7 \times 15 = 190 \text{ cm}$. Thus, the range error of the present system is between 7.5 and 190 cm, depending on the range of the satellite.

*See Lehr (1968b). There is an observed scintillation (energy variation) in the return signal. However, experimental data show that 96% of all returns are not less than 20 dB below the value calculated from the range equation. Thus, the probability that the precision of the range measurement is $h\tau$ is no longer 95% but $0.96 \times 95 = 91\%$.

We have seen that the range accuracy of a laser system is improved if centroid detection rather than threshold detection is used. The accuracy also increases when the intensity of the returned signal is increased. In the next section we look into ways of increasing the signal intensity.

4.2.2 The effect of the intensity of the received signal

The error in the range measurement is inversely proportional to the cube root of the received energy for threshold detection, and to the square root of the received energy for centroid detection. Consequently, this energy should be as large as possible. Possible ways of increasing this energy are discussed below:

A. Increase the laser power. The power of present Q-switched lasers probably cannot be increased significantly without damaging the ruby or the optical components. Modifications of the Q-switched laser combine increased power with reduced pulse duration. These lasers have attractive features for the present application. They are considered in Section 4.2.3.

B. Narrow the transmitted laser beam. Present laser beamwidths are at the limits of visual-tracking errors. Narrower beams require that the satellite direction be predicted. Predictions improve with the accuracy of the field-reduced observations. Such improved observations might be obtained by improving the field reduction of camera observations. And, of course, very accurate field-reduced observations would be available from a worldwide network of laser stations.

C. Narrow the reflected laser beam. The beamwidths of the reflectors on satellites now in orbit are as narrow as velocity aberration will permit. If narrower beams were used, the receiver would have to be moved to a new position relative to the transmitter for each laser pass. And, of course, each displacement between transmitter and receiver would have to be measured to within the desired accuracy of the range measurement. This possibility is not an attractive one.

D. Increase the efficiency of the receiver. The efficiency of the receiver is limited by the aperture of the telescope and the quantum efficiency of the photomultiplier. Possible improvements appear to be either small or costly.

E. Eliminate the target scintillation. This scintillation is probably due to the fact that the retroreflector is an array of hundreds of corner cubes. Other reflectors would be inefficient, or heavy, or not capable of compensating for velocity aberration.

4.2.3 The effect of a shorter pulse

The previous discussion has shown how the precision of the range measurement might be made to correspond to a small fraction of the pulse duration. But the possibility of decreasing the pulse duration itself was not considered. New techniques are becoming available for the generation of pulses much shorter than the approximately 10-nsec lower limit of the Q-switched laser. These techniques have been used successfully in laboratories and might be expected eventually to be applicable under field conditions. For the same resolution, less energy needs to be transmitted when the pulse is short than when it is long. Consequently, present laboratory models of these lasers are attractive in spite of the fact that energies are reduced substantially.

The use of cascaded amplifiers is particularly fascinating, because they increase the energy and reduce the pulse duration simultaneously. In other words, they act to increase the precision of the range measurement in two ways at the same time.

The pulse duration of a Q-switched laser can be made smaller by cavity dumping (Hook, Dishington, and Hilberg, 1966) or mode locking (Heynau and Foster, 1968). In cavity dumping, the transmission of one of the mirrors that comprise the laser's resonant cavity is suddenly increased to unity at the peak of the Q-switched pulse. Then all the stored energy in the cavity is

emitted in twice the time it takes light to traverse the cavity. This mode of operation produces a pulse whose length is of the order of 1 nsec and whose energy is of the order of 1 J.

The mode-locked laser uses an expander (a bleachable dye) within the cavity. The expander attenuates the low-amplitude portions of the laser pulse more than it does the high-amplitude portions. As the light wave is reflected back and forth within the cavity, the pulse narrows. This narrowing keeps up until the pulse's spectral width has increased to that of the cavity. Pulses as short as 10 psec (equivalent to a distance of 1.5 mm) have been produced in this manner. These pulses were produced with glass rods containing neodymium ions as the active element. The wide spectral width of the glass laser is particularly favorable to the generation of short pulses, but the 1.06- μm radiation is too low to be detected with a photomultiplier. Consequently, the transmitted radiation must go through a frequency doubler, or a solid-state detector must be used in the receiver. Another problem arises from the fact that a train of short pulses is generated, with the interpulse interval of the order of nanoseconds. In order to avoid ambiguity in the range measurements, a single one of these pulses must be selected for transmission.

It has been stated (Heynau and Foster, 1968) that the generation of single 1-J pulses at durations of 100 psec is possible. A 100-psec pulse corresponds to a precision in range of 0.6 cm when only one photoelectron is generated by the receiver. How does this precision compare to that of a 10-nsec, 10-J Q-switched laser? For comparable accuracy, the h value for the Q-switched laser would be 10^{-3} . Consequently, even with centroid detection, the received signal would have to be 10^6 times larger for the 10-nsec laser than for the 100-psec laser. But the transmitted energy is only 10 times larger. Consequently, the mode-locked laser has a considerably greater capability for range precision than does the higher energy 10-nsec laser.

4.2.4 The effect of the precision of the interval timing

The last section showed that the potential exists for the construction of picosecond lasers that can lead to range precisions of millimeters if enough amplifiers can be cascaded to achieve the required energy output. We now ask whether a receiving system can be built to realize in practice the precision that could be available when such a laser system is used.

We consider detection with a photomultiplier (Ross, 1966, pp. 140-163; Lodge, Muff, Owen, and Smout, 1968), the device that at present appears to be most suitable when the precise timing of short pulses of visible radiation is required. The time dispersion of the photoemission is about 3 psec, and the time dispersion of the secondary emission is about 6 psec. These are the fundamental limits to the precision of a photomultiplier. But these limits are never reached in practice. The problem is that the current emitted by the photocathode must be amplified by a factor of 10^6 to 10^8 . This amplification cannot be achieved in practice without a large increase in the time dispersion. Difficulties arise from the fact that the electrons are emitted from the photocathode with varying energies and directions. The amplifier can be designed for only single values of these parameters. It is also true that perfectly uniform electric and magnetic focusing fields cannot be achieved within a finite volume. This fact also introduces limitations in the design of the amplifier.

An electrostatic photomultiplier (the one commonly in use at present) has a frequency response that extends to about 100 MHz. This response corresponds to a resolution of 10 nsec (150 cm). Newer photomultipliers whose amplifiers use crossed electric and magnetic fields are reported to have resolutions of about 300 psec (4.5 cm). Commercial time-interval counters and oscilloscopes have about the same limiting resolutions.

4.2.5 The effect of the atmospheric correction

The correction for atmospheric effects was described above in connection with the laser system now operating at the Mt. Hopkins Observatory. This correction is based on the use of measured values of the temperature and barometric pressure. The associated error is 5 cm. For a more precise correction, a better estimate of the average index of refraction along the path of the laser beam must be made. Such an estimate is possible if range measurements are made at two wavelengths simultaneously (Owens, 1967; Fowler, 1968). This method works because the shape of the dispersion curve (index of refraction versus wavelength) does not change with the state of the atmosphere: All points on the curve go up and down proportionately with atmospheric changes. Consequently, if the atmospheric correction at wavelength λ_1 is S_1 , and at λ_2 is S_2 , we have the relation

$$S_2 = \frac{S_2 - S_1}{A} , \quad (18)$$

where A is a constant that can be calculated from the known dispersion curve of air at standard temperature and pressure. The quantity $S_2 - S_1$ is just the difference between the two range measurements. The corrected range is $R_2 - S_2$, where R_2 is the measured range at λ_2 , and S_2 comes from equation (18). A difficulty with this method comes about because A is less than 1. For $\lambda_1 = 694 \mu\text{m}$ and $\lambda_2 = 347 \mu\text{m}$, A is 0.11. Consequently, the error in the value of S_2 computed from equation (18) is 10 times the error in $S_2 - S_1$. In other words, the range must be measured to within an error that is one-tenth of that remaining after the atmospheric correction is applied. For this reason, the two-wavelength procedure will be useful only when an instrumental precision better than 5 mm is achieved.

4.2.6 The effect of the stability of the system's delay

The laser system measures the time from the transmission of a laser pulse until its return from the satellite. This time is not only that associated with the two-way travel time through the atmosphere, but also that of system

delays, both optical and electrical. These, lumped together, are measured by ranging on a nearby reflector at a known distance. The delay correction is obtained by comparing the known range and the measured range. The delays can be of the order of 100 nsec (15 cm). For precise ranging, it is obviously true that this delay must be accurately determined. It also must remain stable between successive calibrations. So many factors are involved in the stability of the delay that it appears impossible to calculate a meaningful estimate. It is something that will have to be determined empirically. But its importance may turn out to be very great. In fact, it may well limit the precision of a laser system.

4.2.7 The effect of the accuracy in determining the epoch of the measurement

In any consideration of the accuracy of the range measurement, the uncertainty in the time of the observation (i. e., the uncertainty in the epoch) must be consistent with the uncertainty in the measurement of the distance. A satellite in a circular orbit 1 Mm (620 miles) from the earth has a velocity of 7.4 km sec^{-1} . At this velocity, its range changes less than 1 m in 100 μsec . This is approximately the accuracy to which the worldwide synchronization of SAO clocks is currently maintained. For an accuracy of 1 cm, timekeeping to 1 μsec will be required. This accuracy can be achieved at present (Shapiro, 1965). It would even be possible to keep time to better than 1 nsec if independent-clock radio interferometry were used (Gold, 1967).

5. ATMOSPHERIC AND IONOSPHERIC INFLUENCES

M. D. Grossi

5.1 Introduction

The accuracy of geodetic measurements made on the earth's surface by stations of a network receiving electromagnetic emissions from artificial satellites and stellar sources will ultimately be limited by the turbulence and unpredictability (Tatarski, 1961) of the ionosphere and atmosphere through which the received waves are propagated.

In the absence of atmospheric and ionospheric degrading effects, recent progress in instrumentation technology would allow geodetic measurements with station positional errors of a few centimeters. However, no fully established method yet exists that is capable of measuring these atmospheric and ionospheric effects with the necessary precision, although some avenues of approaching the problem show definite promise.

We review here the refraction effects of the ionized and neutral atmosphere upon microwave propagation, and possible methods of atmospheric and ionospheric probing applicable to geodetic measurements.

5.2 Atmospheric and Ionospheric Differential Phase Path Length and Its Fluctuations

Differential phase path length is defined as the difference between geometric and electromagnetic wave path length between two preestablished points. With μ the index of refraction of the medium, the differential phase path length between two points A and B of the medium is

$$\int_A^B ds - \int_A^B \mu ds = \int_A^B (1 - \mu) ds \quad .$$

In the ionosphere, at microwave frequencies, the phase velocity is larger than the velocity of light and μ is smaller than 1. The differential phase path length is therefore a positive number. In the atmosphere, μ is larger than 1 and the differential phase path length is a negative number. For a microwave radio path that traverses both the atmosphere and the ionosphere, the overall effect is a negative differential phase path length.

The data in Table 8 have been obtained from the models in Figures 13 and 14 and from a ray-tracing computer program. In the table, the ground station was at midlatitude, with local time at midmorning. The numbers in the table therefore represent typical values of the so-called cumulative refractive "bias" of the atmosphere and ionosphere at the two tabulated frequencies.

Table 8. Surface-to-free-space differential phase path length

	5000 MHz	10,000 MHz
Looking at the zenith	- 1.956 m	- 2.263 m
45° away from zenith	- 2.848 m	- 3.218 m
Grazing the horizon	-85.67 m	-86.65 m

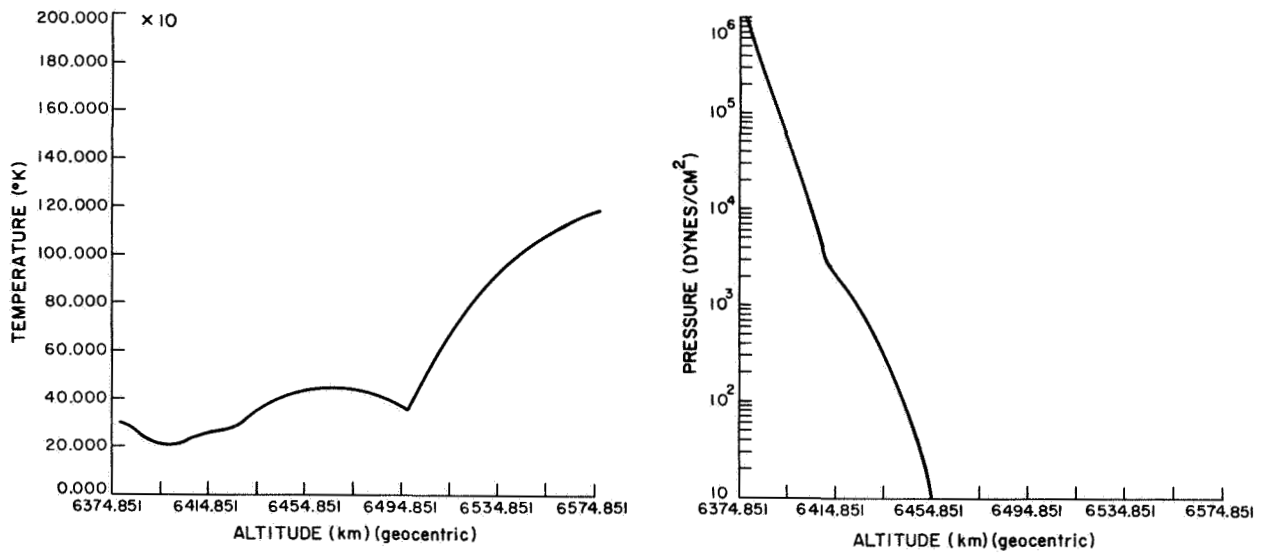


Figure 13. Spherically symmetric model of dry-air atmosphere (Harrington, Goff, Grossi, and Langworthy, 1966); $\mu = 1 + 79 \times 10^{-6} p/T$ (p in millibars, T in °K).

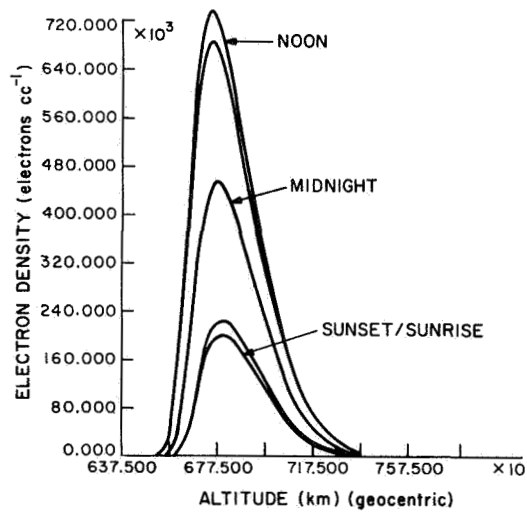


Figure 14. Ionospheric electron-density profile (Harrington et al., 1966).

It can be seen that, if uncorrected, the refractive effects of the atmosphere and ionosphere would hardly be tolerable for geodetic applications of satellite or radio-star observations. Large-scale, slow-dynamics atmospheric and ionospheric models have been developed, however, that could be updated to the time and place of the observations. By use of these models, a large part of the refractive bias can be removed from the measurements. Essentially, this is done by reconstructing by ray tracing the refractive contributions to the radio path length of the model ionosphere and atmosphere. An unrecoverable bias error still remains, however, owing to the lack of realism of the models; according to R. K. Crane (1967, private communication), this residual error amounts to about 15 cm (when looking approximately 45° away from the zenith) at the two frequencies of Table 8 and is almost totally due to atmospheric effects.

For the effect of random fluctuations, a ray-tracing analysis has been carried out by Cunningham and Crane (1965). They evaluated the refractive effect of selected turbulent cross sections in the lower atmosphere crossed by a radio path that was moving angularly. The rms value of the fluctuations in phase path length was computed to be approximately 14 cm. Again, at the two frequencies of Table 8, the ionospheric contributions to the random fluctuations can be considered negligible.

However, microwave frequencies of approximately 1000 MHz or lower, ionospheric corrections (bias and fluctuations) cannot be disregarded.

On the experimental side, Thompson, Janes, and Kirkpatrick (1960) of the National Bureau of Standards conducted measurements of radio path length at 9400 MHz over a 15.5-mile path on Maui, Hawaii (7° above horizon), and over a 9.5-mile path near Boulder, Colorado (11' above horizon). Their measurements indicate fluctuations of phase path lengths close to Crane's analytical results. Figure 15 shows a typical record, with a "trend" between November 6 and November 9, 1956 (the phase path length was decreasing approximately 12 cm per day), and with superimposed fluctuations of approximately 15-cm rms. The spectrum of these measurements is given in Figure 16.

Figures 17 and 18 show examples of path-length fluctuations for the Mesa Verde link. Figure 19 provides the spectrum of the fluctuations of the example given in Figure 18. The width of the spectra in the frequency domain and their spread along the y axis at each frequency (which suggests that the process involved is not stationary) point out the unreliability of large-scale, slow-dynamics models for correction of the bias and of the random fluctuations.

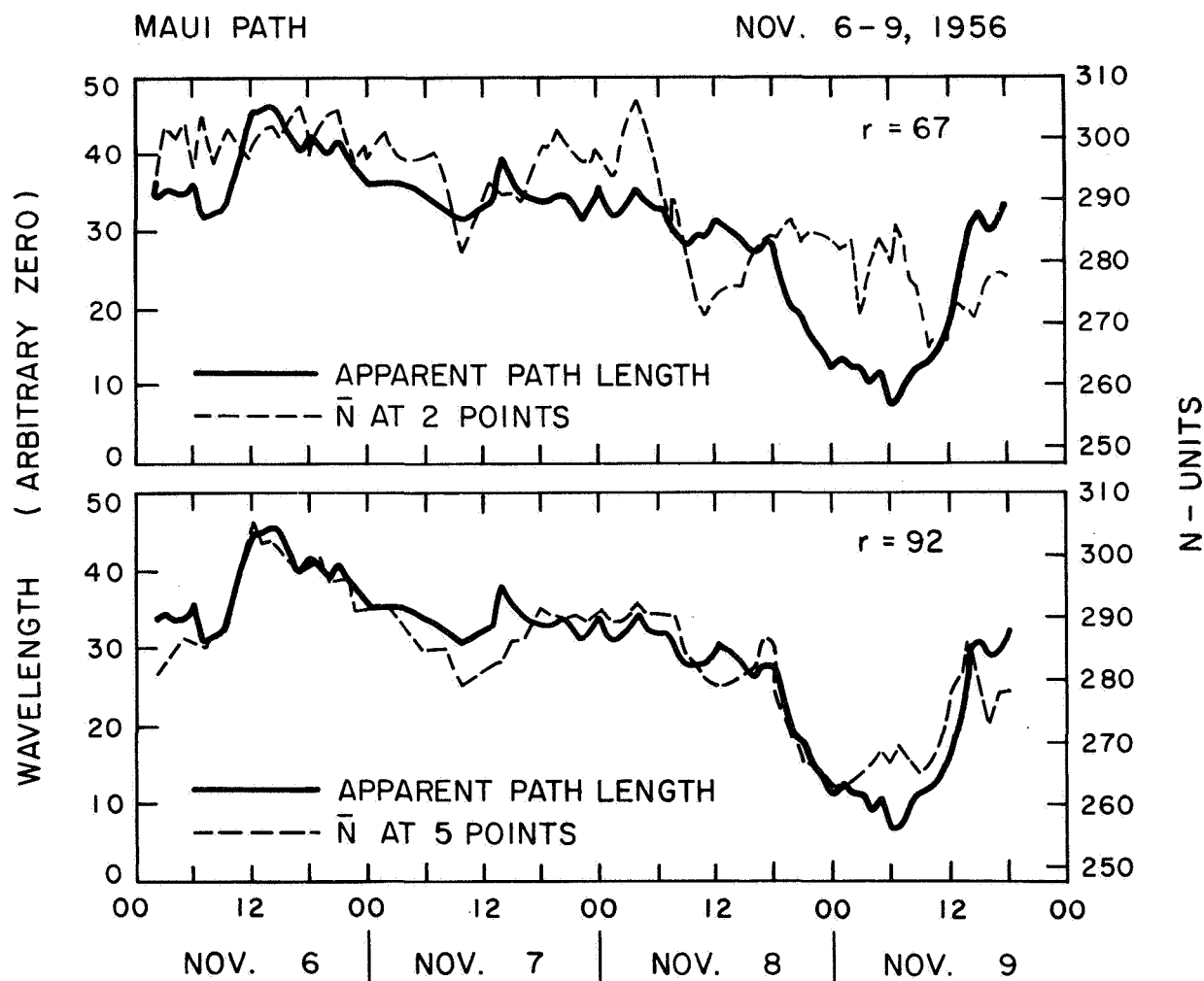


Figure 15. Long-term changes in electrical path length and refractivity, Maui path (November 6-9, 1956). From Thompson et al. (1960).

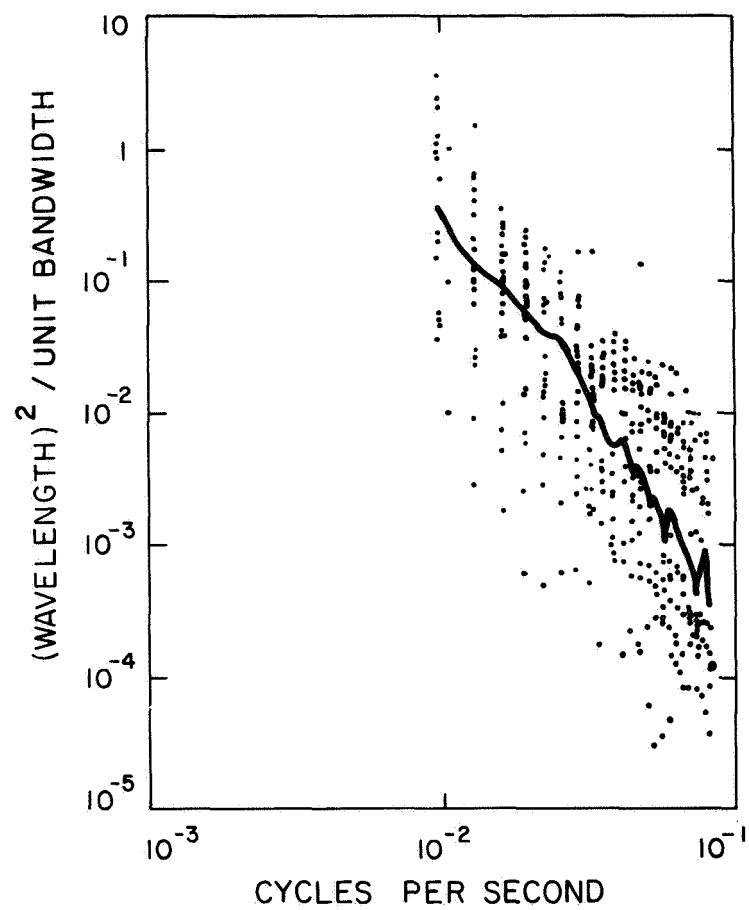


Figure 16. Spectrum of apparent path-length variations, Maui path (November 6-9, 1956). From Thompson et al. (1960)

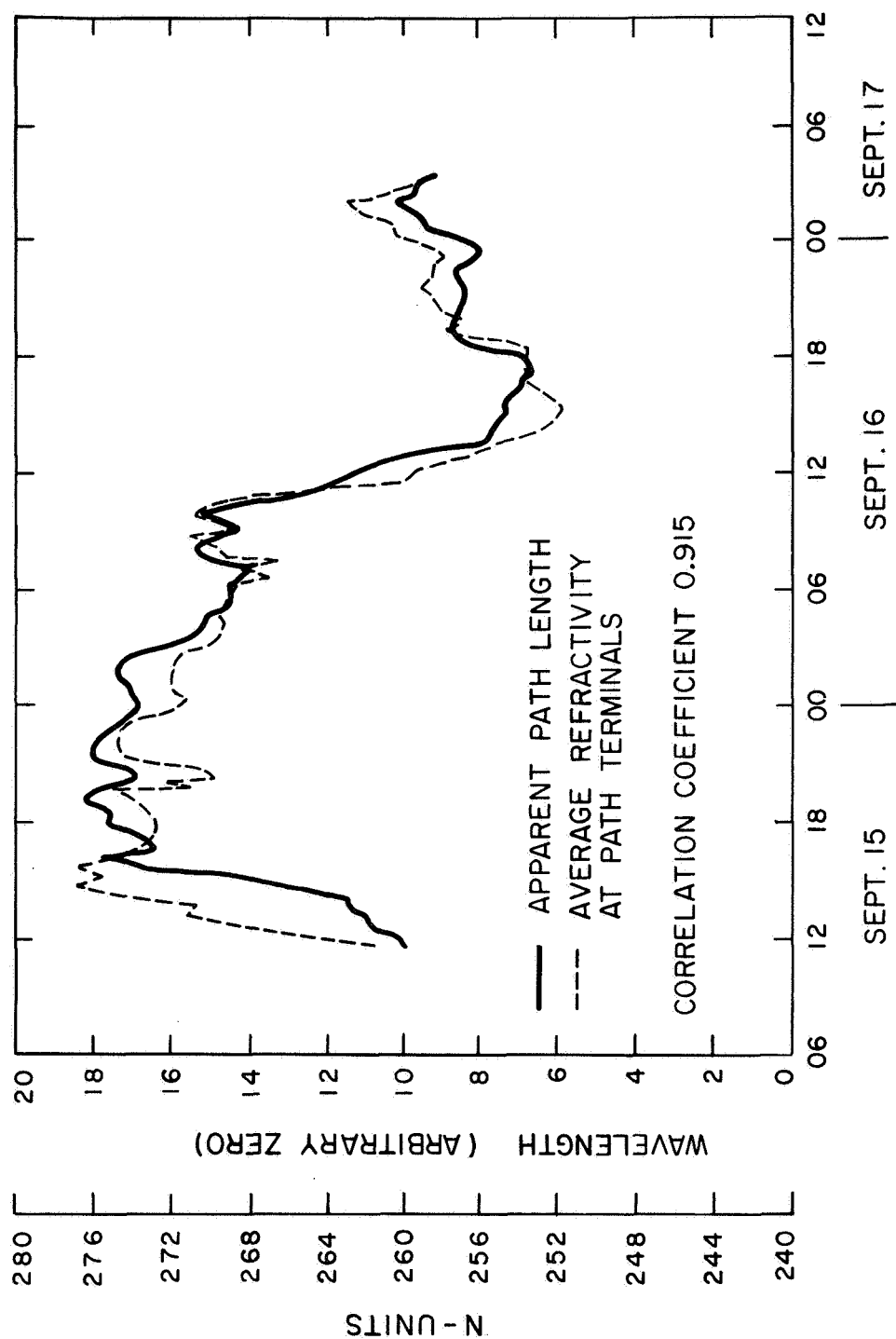
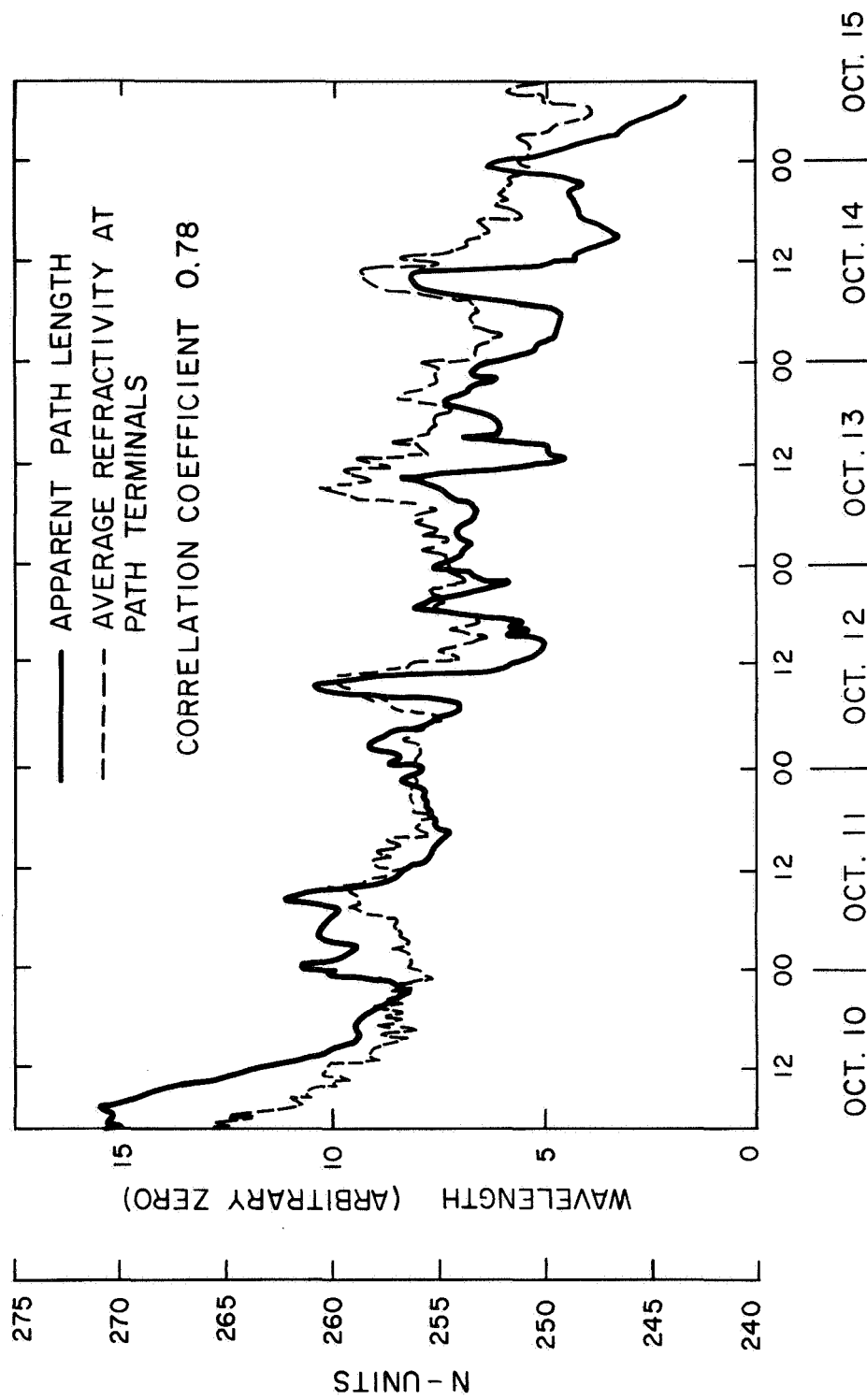


Figure 17. Long-term changes in electrical path length and refractivity, Green Mountain Mesa -- Table Mesa path (September 15-17, 1958). From Thompson et al. (1960).



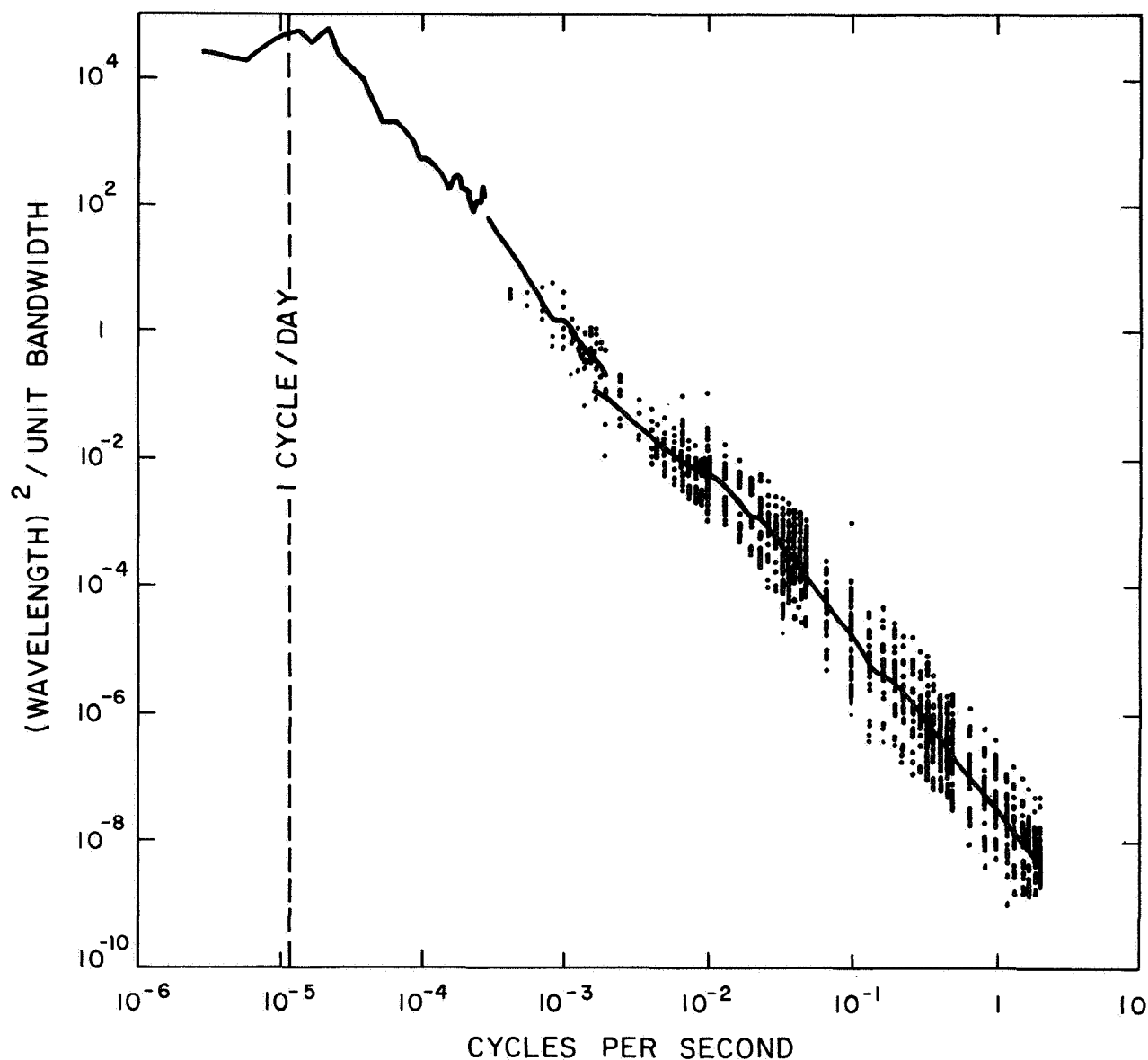


Figure 19. Spectrum of apparent path-length variations, Green Mountain Mesa – Table Mesa path (October 10-15, 1958). From Thompson et al. (1960).

Measurements performed with balloon-borne microwave refractometers have since shown that in fact the width of the autocorrelation function in the time domain of a vertical profile is on the order of 15 min and that profiles measured simultaneously along two verticals only 2 miles apart are already fully uncorrelated.

Four-dimensional refractivity models capable of describing the medium with this degree of sophistication are not yet available, and radiosonde or refractometer probings (Bean and Dutton, 1966) performed with adequate spatial and temporal resolution around the geodetic station do not appear practical.

Instead, the measurement of the columnar atmospheric and ionospheric refractivity could and should be performed along the same radio path involved in the geodetic measurements and should be repeated as frequently as required by the temporal variations of the medium or by the angular motion of the direction of observation.

When satellites are being observed, radio emissions from on-board transmitters can be designed to provide the required information on the columnar refractivity along the path. When geodetic measurements are performed by observing a radio star, the ground terminal should be equipped with radar-type sensors in order to perform monostatically columnar-refractivity measurements along the path.

5.3 Columnar-Refractivity Measurements in the Direction of Observation

When satellite observations are involved, ionospheric columnar refractivity can be measured as follows: The satellite radiates two coherent frequencies in the microwave spectrum near the value of the observation frequency. The differential doppler observed on the ground provides the measurement of the rate of change versus the time of the columnar ionospheric refractivity (de Mendonca, 1963). With certain restrictions on the motion of the satellite with respect to the ionospheric stratification, the integration in the time domain of the differential doppler provides the columnar ionospheric electron content in the direction of observation. The constant of integration can be obtained by associating to the differential doppler the determination of the Faraday rotation at both frequencies.

For the determination of the columnar refractivity of the neutral atmosphere, a similar approach can be followed, with a space-to-ground link working at wavelengths for which the atmosphere is dispersive. Thayer (1967) has developed a differential group-delay scheme that uses two frequencies in the optical band (3660 Å and 6328 Å). The method requires three frequencies when the location of the ground station is such that the refractive effects of water vapor in the atmosphere are significant.

This bistatic probing of the ionosphere and atmosphere is expected to yield rms errors in differential phase path-length measurements of a few centimeters for zenith angles smaller than 45° and for observing frequencies in the microwave band.

When radio astronomical observations of discrete sources with small angular size are the basis of the geodetic measurements, we have to resort to radar techniques (Figure 20) for the measurements of the total columnar refractivity along the direction of observation.

Microwave radar measurements of the incoherent backscatter from the electrons of the ionosphere (Evans, 1964) allow the direct plotting of the electron-density profile along the wanted direction. This technique is currently used at vertical incidence in conjunction with an ionosounder. The incoherent-backscatter radar provides the profile in arbitrary density units, while the sounder "normalizes" it by determining the electron content at the peak of the F_2 layer ($F_2 N_{\max}$). In the configuration of Figure 20, the radar could be calibrated to provide directly in electrons per cubic centimeter the ionospheric profile along the direction of the radio astronomical observations. The large-size parabolic reflector and part of the receiver used for the radio astronomical observations could be shared with the ionospheric radar.

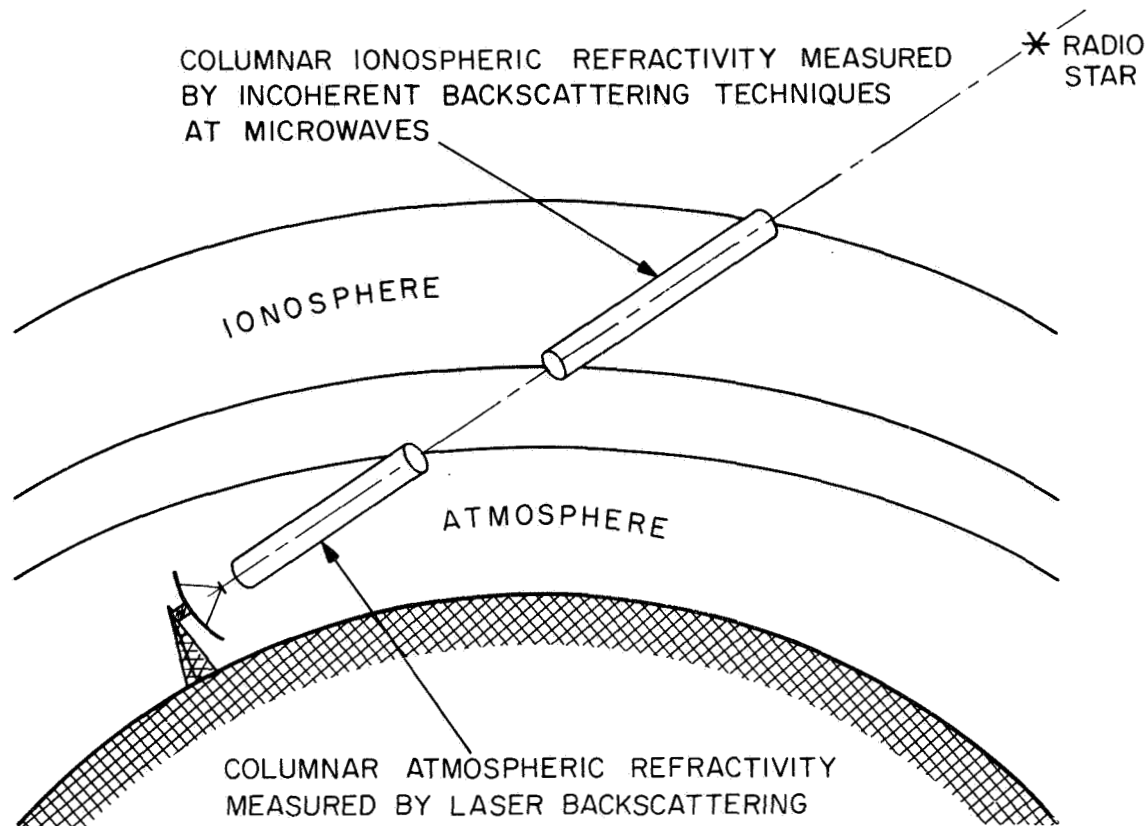


Figure 20. Proposed radar arrangement for the monostatic measurement of the atmospheric and ionospheric columnar refractivity in the wanted direction of observation.

To measure monostatically the atmospheric columnar refractivity, a pulsed ruby laser could be used (Fiocco, 1965) and the backscatter intensity at the various ranges could be interpreted in terms of the number density of the atmosphere. The measured number density could then be converted into atmospheric refractivity.

5.4 Conclusions and Recommendations

The measurement by long-base interferometers of angles of arrival of microwave emissions from radio stars or satellites with rms errors better than $1/50 \mu\text{rad}$ requires the determination of atmospheric and ionospheric differential phase path lengths above each of the interferometer terminals with rms errors better than 5 to 10 cm.

This goal does not appear unreasonable if the total columnar refractivity can be measured along the radio path of the observation and at instants of time that bracket closely the observation time.

The case of satellite observations with accuracy that meets the geodetic requirements is being treated extensively in the literature, and the best course of action for SAO seems to be to keep abreast of the latest developments and to update its computational tools to include recent procedures of ionospheric and atmospheric refractivity corrections.

For radio astronomy, on the other hand, a specific research effort is required. Analytical and experimental efforts must be performed to ascertain the feasibility and practicality and the accuracy limitations of the laser probing of the atmospheric refractivity and the adaptability of absolute-calibration incoherent-backscattering radar techniques to the determination of the ionospheric correction.

6. MODELING THE SYSTEM

S. Ross

Formulating and programming a model of the system for measuring continental drift becomes worthwhile when the number of possible configurations and the variety of parametric choices within each concept are large enough to warrant exercising the model repeatedly over a number of possible alternatives. Modeling the system is especially worthwhile when its elements are so complex that it becomes unfeasible to study their interrelationships with a simple analytical approach.

Let us first examine some of the most fundamental questions that govern the establishment of a program for measuring small crustal motions of the earth.

A. Which approach, laser or interferometer, should be used? Or should it be a combination of both, and if so, what sort of combination?

B. If an interferometer is used, should it track stellar or satellite sources, or both?

C. If satellites are used, how many should there be and what should their orbital characteristics be?

D. Where should observing stations be placed?

E. What will the program costs amount to, considering expenditures for equipment, network operation, and (possibly) development and purchase of spacecraft?

F. How "effective" will any proposed system be, if not in absolute terms, then at least relative to other reasonable alternatives?

Of course, answers to these questions will be highly subjective, and no model can be expected to settle them directly. What the model can do, however, is provide quantitative information from which subjective decisions can be made with some reasonable sense of assurance that the best available data and theories have been used in formulating the policies. For example, the number of satellites and their deployment will depend in part on where the observation stations are located and in part on the costs involved in using that number of satellites with that particular station configuration. Station and network maintenance costs depend in large measure on how often the various stations are used and on what their duty cycles turn out to be in each instance. These factors, in turn, depend on which object is being tracked and where it is at what times. They also depend on some statistical estimate of the climatic conditions that affect viewing at the different station locations and on the characteristics of the equipment at the stations. Perhaps the hardest question to answer is the sixth, because it is far from obvious what constitutes a valid criterion for effectiveness.

What is clear is that in order to make the major policy conclusions meaningful in a practical sense, we must consider the interplay among all the diverse aspects of the problem—operations, performance, and cost. This approach by no means obviates the necessity for carrying out studies of any single aspect in depth; on the contrary, detailed studies of individual parts done in relative isolation from the effects of other parts of the system are still necessary to point out reasonable ranges of parametric values to use in the broader studies. If such preliminary analysis cannot serve to restrict the number of cases to be covered later, it will generally not be feasible to conduct the more comprehensive studies.

The next point to observe is that many basically different versions can be constructed for almost any submodel, and that they may differ from one another either in the essential nature of approach or in the degree of detail used to formulate them. Each has its place; the more cursory are useful in preliminary studies, and the more exact (usually slower running) are useful in advanced phases of system definition. An orbit-coverage program

based on statistical averaging of the satellite's position is very useful in analyzing problems in preliminary phases, while deterministic models must be applied to examine in greater detail what occurs during some specific portion of a given flight.

All this is well known to anyone who has had experience in analyzing large systems. What is worth noting, however, is that this system, like most systems, can be conveniently broken down into interacting elements, for each of which many available models can usually be found. Our approach has been first to identify in a preliminary way the basic elements that enter into the system, together with the pattern of flow of information between these elements. Next, we have begun to survey what is available in the way of existing computer models for the various elements and, where appropriate, to identify the data bases that will serve the elements we collect. Finally, we have converted and checked out on the CDC 6400 an existing executive system that is capable of incorporating within its framework the various program modules mentioned above. The executive automatically sets up linkages among these program modules to allow them to be used as an ensemble in any specific problem. The operational characteristics of the executive system are described by Gold and Ross (1968).

Figure 21 depicts a preliminary overall model for evaluating the performance of candidate measurement systems. The analyst specifies a proposed network of stations and STATIONS computes the orientations and lengths of the set of base lines that connect each pair of stations. GEOLOGY contains digitized data on the locations of fault lines, ocean rises, trenches, etc. and on the presumed rates of drift relative to them. It calculates which of the network base lines will cross these zones, and at what relative heading (azimuth) angles. This information is useful later in the assessment of the effectiveness (EFFECT) of the configuration being studied.

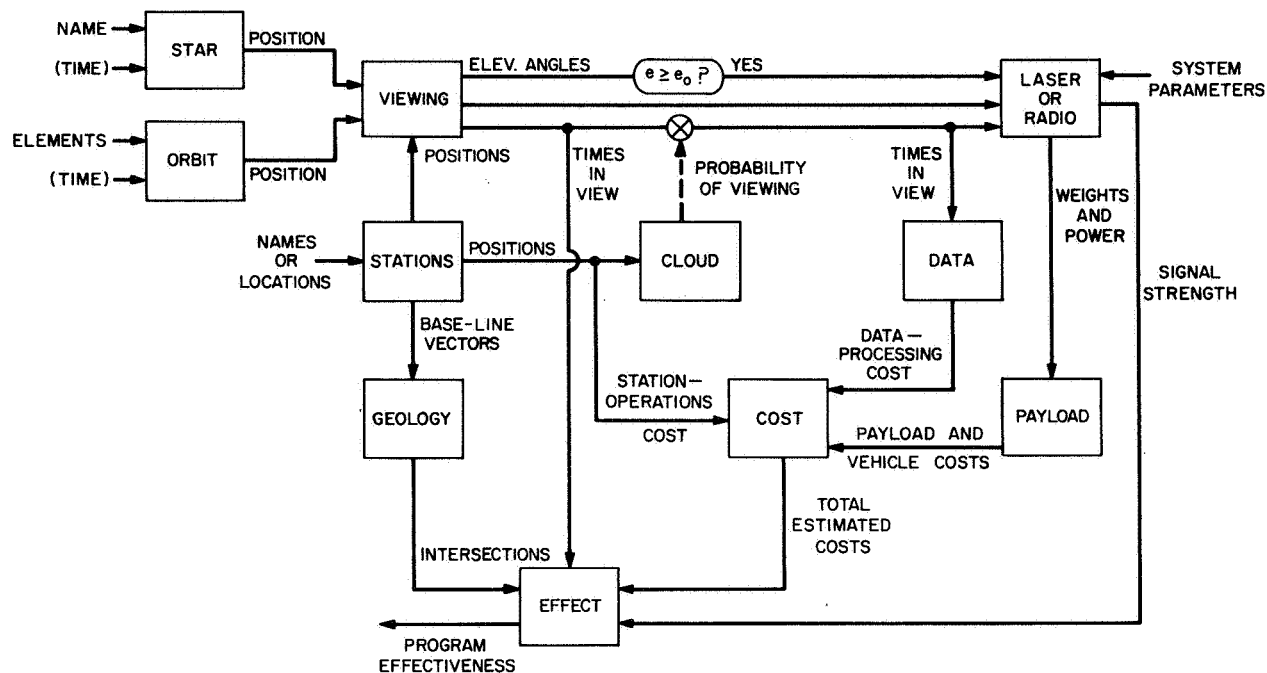


Figure 21. Preliminary flow chart for total position measurement system.

STAR and ORBIT are sets of programs that generate positional information for radio sources and for satellites, respectively. Currently, we plan to have two sets of programs for each case: One set embodies a probabilistic approach, to furnish statistical information on positional distributions of the source averaged over long periods of time; the other involves a deterministic approach, to give specific positions as a function of the viewing time specified as input. The choice of which of the four is to be used is exercised through the executive system. If one of the ORBIT programs is used, its output in terms of either satellite positions or distribution of positions is combined in VIEWING with the station-position vectors that are output from STATIONS, to determine the elevation (i. e., pointing) angles for the ground equipment and to test whether minimum

elevation constraints have been violated. If STAR happens to be selected, analogous calculations are performed for the radio source. In both cases, times in view are also calculated, as well as slant ranges to the satellite, where appropriate.

CLOUD is a program that contains digitized statistics on cloud cover over various regions of the world during different parts of the day, for each season. Its output is combined in PCT with the elevation angles obtained from VIEWING, to give distributions of probabilities of viewing the object from the stations specified within the selected time periods. It may be possible to add a section that estimates signal attenuation and distortion effects due to the atmosphere and ionosphere.

LASER and RADIO estimate the performance of the ground-based and space-borne elements of the tracking system, outputting data on weights and power levels for the space-borne equipment, as well as the expected characteristics (signal-to-noise ratio, flux levels, etc.) of the incoming signal to the receiving antenna. If a spacecraft is to be used, PAYLOAD estimates the size and weight of the payload package and support hardware (e.g., power supplies). COST estimates costs for the configuration from time-in-view information, payload-weight and power information, and (perhaps) the volume and rates of data to be processed.

Because of the modular approach, any element can be replaced when a more suitable model is provided, without disturbing the rest of the system. No firm plans have been made yet for the format of EFFECT, pending more experience with the system.

Programs exist for CLOUD (Sherr et al., 1968), for the deterministic ORBIT, and for portions of RADIO.* Digitized data exist on the locations and sizes of large radio telescopes, on the Baker-Nunn and SAO laser tracking networks,

*Any module can itself be decomposed into submodules.

and on the centers of rotation and relative drift rates of continental blocks (Le Pichon, 1968). Preliminary information has been obtained on some aspects of tracking-network expenditures and on laser and radio-astronomy system costs. SAO has determined that it is possible to formulate at least an empirical model of the laser system performance (Lehr, 1968a).

Probabilistic models for ORBIT and STAR will have to be developed. If the cylindrical equal-probability transformation

$$x = \sin^{-1} \left(\frac{\sin L}{\sin i} \right) ,$$

$$y = \lambda ,$$

where L is the instantaneous longitude, λ is the latitude, and i is the inclination, is applied to the satellite orbit (or the star's "orbit" relative to a stationary earth), equal areas in the x - y plane will be traversed by the object in equal times. Expressed alternatively, the probability of finding the satellite or star within any closed curve of visibility will be equal to the ratio of the area of the curve to the total area lying between latitudes of $\pm i$. The area contained within the visibility curve (cf. Section 3) is determined by numerical integration using Stokes' theorem. We have programmed such a module and are checking it out on the CDC 6400.

It is intended to write some plotting programs for the SC 4020 (see the discussion of automatic plotting in Gold and Ross, 1968) that will present stereographic and plane-projected views of the base lines used in any problem, overlaid on digitally drawn maps of the earth's surface with fault zones indicated. We have acquired almost all the data base needed for such projections.

7. REFERENCES

BARAZANGI, M., and DORMAN, J.

1968. World seismicity map of ESSA Coast and Geodetic Survey epicenter data for 1961-1967. Bull. Seismol. Soc. Am., vol. 58, in press.

BARE, C., CLARK, B. G., KELLERMANN, K. I., COHEN, M. H., and JAUNCEY, D. L.

1967. Interferometer experiment with independent local oscillators. Science, vol. 157, no. 3785, pp. 189-191.

BEAN, B. R., and DUTTON, E. J.

1966. Radio meteorology. NBS Monograph 92, 435 pp.

BENDER, P. L., ALLEY, C. O., CURRIE, D. G., and FALLER, J. E.

1968. Satellite geodesy using laser range measurements only. Journ. Geophys. Res., vol. 73, pp. 5353-5358.

BLACKMAN, R. B., and TUKEY, J. W.

1958. The Measurement of Power Spectra, from the Point of View of Communications Engineering. Dover Publ., New York, 190 pp.

CUNNINGHAM, R. M., and CRANE, R. K.

1965. Tracking errors due to selected atmospheric cross-sections. AFCRL Tech. Rep. P 59301, May.

DE MENDONCA, F.

1963. Ionospheric studies with the differential doppler technique. In Radio Astronomical and Satellite Studies of the Atmosphere, ed. by J. Aarons, North-Holland Publ. Co., Amsterdam, pp. 289-312.

EVANS, J. V.

1964. Incoherent backscatter studies of the ionosphere at Milestone Hill. In Electron Density Distribution in Ionosphere and Exosphere, ed. by E. Thrane, North-Holland Publ. Co., Amsterdam, pp. 266-304.

FELLER, W.

1968. An Introduction to Probability Theory and Its Application. Vol. I, 3rd ed., J. Wiley and Sons, New York, p. 228.

FIOCCO, G.

1965. A summary of research with optical radars. 20th Annual ISA Conference, Los Angeles, California, October 4-7, preprint no. 40.1-3-65.

FOWLER, R. A.

1968. Earthquake prediction from laser surveying. NASA SP-5042, 32 pp.

GOLD, T.

1967. Radio method for the precise measurement of the rotation period of the earth. Science, vol. 157, no. 3786 pp. 302-304.

GOLD, R., and ROSS, S.

1968. Automated mission analysis using a parametric sensitivity executive program. Presented at AAS/AIAA Astrodynamics Specialist Conference, Jackson, Wyoming, September 3-5, 6 pp.

GOODMAN, J. W.

1965. Some effects of target-induced scintillation on optical radar performance. Proc. IEEE, vol. 53, pp. 1688-1700.

HARRINGTON, J. V., GOFF, R. J., GROSSI, M. D., and LANGWORTHY, B. M.

1966. Electromagnetic measurements of planetary atmospheres and ionospheres by an orbital pair. Raytheon Tech. Rep. FR-66-71, March 7.

HEYNAU, H. A., and FOSTER, M. C.

1968. Single-subnanosecond laser pulse generation and amplification: the second generation of Q-switched lasers. Laser Focus, August, pp. 20-24.

HOOK, W. R., DISHINGTON, R. H., and HILBERG, R. P.

1966. Laser cavity dumping using time variable reflection. Appl. Phys. Lett., vol. 9, pp. 125-127.

ISACKS, B., OLIVER, J., and SYKES, L. R.

1968. Seismology and the new global tectonics. Journ. Geophys. Res., vol. 73, pp. 5855-5899.

JENNISON, R. C.

1966. Introduction to Radio Astronomy. Philosophical Library Publ. Co., New York, 160 pp.

LAUGHTON, A. S.

1966. The Gulf of Aden. Phil. Trans. Roy. Soc. London, A, vol. 259, pp. 150-171.

LEHR, C. G.

- 1968a. Atmospheric correction to laser observation. Smithsonian Astrophys. Obs. Laser Report No. 3, 3 pp.
1968b. The energy distribution of the received signals in a ruby-laser satellite-ranging system. Smithsonian Astrophys. Obs. Laser Report No. 2, 35 pp.

LE PICHON, X.

1968. Sea-floor spreading and continental drift. Journ. Geophys. Res., vol. 73, pp. 3661-3697.

LODGE, J. A., MUFF, P., OWEN, R. W., and SMOUT, D.

1968. A cross-field photomultiplier with sub-nanosecond rise time suitable for nuclear studies. IEEE Trans., vol. NS-15, pp. 491-495.

MARVIN, U. B.

1966. Continental drift. In Scientific horizons from satellite tracking, ed. by C. A. Lundquist and H. D. Friedman, Smithsonian Astrophys. Obs. Spec. Rep. No. 236, pp. 31-74.

MORAN, J. M.

1968. Interferometric observations of galactic OH emission. Ph. D. Thesis, MIT.

MORAN, J. M., CROWTHER, P. P., BURKE, B. F., BARRETT, A. H., ROGERS, A. E. E., BALL, J. A., CARTER, J. C., and BARE, C. C.

1967. Spectral line interferometry with independent time standards at stations separated by 845 kilometers. Science, vol. 157, no. 3789, pp. 676-677.

MORGAN, W. J.

1968. Rises, trenches, great faults, and crustal blocks. Journ. Geophys. Res., vol. 73, pp. 1959-1982.

MORRISON, T. S., and ACKERMAN, S.

1967. Analysis of a programmed multipulse laser range measurement system. Appl. Opt., vol. 6, pp. 1725-1727.

OWENS, J. C.

1967. The use of atmospheric dispersion in optical distance measurements. Presented at 14th General Assembly, International Union of Geodesy and Geophysics, Lucerne, Switzerland, September 25 to October 7.

OWENS, J. C., and BENDER, P. L.

1967. Multiple wavelength optical distance measurement. In Electromagnetic Distance Measurement, Hilger and Watts, Ltd., pp. 159-164.

ROSS, M.

1966. Laser Receivers. J. Wiley and Sons, New York, 405 pp.

SHAPIRO, L. D.

1965. Loran-C timing. Frequency, March-April, pp. 32-37.
1968. Time synchronization from Loran-C. IEEE Spectrum, August, pp. 46-55.

SHERR, P. E., et al.

1968. World-wide cloud cover distributions for use in computer simulations. Allied Research Assoc., Inc., Concord, Massachusetts, January.

SYKES, L. R.

1968. Seismological evidence for transform faults, sea-floor spreading, and continental drift. In Proceedings of NASA Symposium, History of the Earth's Crust, ed. by R. A. Phinney, Princeton Univ. Press, Princeton, in press.

TATARSKI, V. I.

1961. Wave Propagation in a Turbulent Medium. Translated by R. A. Silverman, McGraw Hill, New York, 285 pp.

THAYER, G. D.

1967. Atmospheric effects on multiple frequency range measurements. ESSA Tech. Rep. IER 56-ITSA53, October.

THOMPSON, M. C., Jr., JANES, H. B., and KIRKPATRICK, A. W.

1960. An analysis of time variations in tropospheric refractive index and apparent radio path length. Journ. Geophys. Res., vol. 65, pp. 193-201.